Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles

Yuta Koike$^{1,2}$

Received: 23 August 2019 / Accepted: 6 October 2020 / Published online: 26 October 2020 © Japanese Federation of Statistical Science Associations 2020

Abstract
Let $X_1, \ldots, X_n$ be independent centered random vectors in $\mathbb{R}^d$. This paper shows that, even when $d$ may grow with $n$, the probability $P(n^{-1/2} \sum_{i=1}^n X_i \in A)$ can be approximated by its Gaussian analog uniformly in hyperrectangles $A$ in $\mathbb{R}^d$ as $n \to \infty$ under appropriate moment assumptions, as long as $(\log d)^5/n \to 0$. This improves a result of Chernozhukov et al. (Ann Probab 45:2309–2353, 2017) in terms of the dimension growth condition. When $n^{-1/2} \sum_{i=1}^n X_i$ has a common factor across the components, this condition can be further improved to $(\log d)^3/n \to 0$. The corresponding bootstrap approximation results are also developed. These results serve as a theoretical foundation of simultaneous inference for high-dimensional models.

Keywords  Anti-concentration inequality · Bootstrap · Factor structure · Maxima · Randomized Lindeberg method · Stein kernel

1 Introduction
Let $X = (X_i)_{i=1}^n$ be independent centered random vectors in $\mathbb{R}^d$ and consider the normalized sum:

$$S_n^X = (S_{n,1}^X, \ldots, S_{n,d}^X)^\top := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

We assume that each coordinate of $S_n^X$ has (at least) a finite second moment and write the covariance matrix as $\mathbb{C}_n^X := E[S_n^X (S_n^X)^\top]$. The aim of this paper is to approximate $S_n^X$ by its Gaussian analog $Z_n^X$ in law, where $Z_n^X = (Z_{n,1}^X, \ldots, Z_{n,d}^X)^\top$ denotes a $d$-dimensional centered Gaussian vector with covariance matrix $\mathbb{C}_n^X$. When $n$ tends to infinity...
while $d$ is fixed, such an approximation is commonly formulated as convergence in law. Then it is merely a consequence of a classical multivariate central limit theorem (CLT) under mild regularity assumptions. Nevertheless, in a high-dimensional setting where $d$ grows as $n \to \infty$, the situation is not as simple as above. In such a setting, it is typical that $Z_{n}^{X}$ depends on $n$ and has no limit law as $n \to \infty$, so the standard formulation is no longer meaningful. One possible way to properly formulate the problem is to consider the convergence of some metric between the laws of $S_{n}^{X}$ and $Z_{n}^{X}$. A typical choice of such a metric is the following one:

$$
\rho_{n}(A) := \sup_{A \in \mathcal{A}} \left| P(S_{n}^{X} \in A) - P(Z_{n}^{X} \in A) \right|
$$

where $\mathcal{A}$ is a class of Borel sets in $\mathbb{R}^{d}$. In this regard, investigation of Lyapunov type bounds for $\rho_{n}(\mathcal{A})$ with explicit dimension dependence has some history in the case that $\mathcal{A}$ is the class of all convex Borel sets in $\mathbb{R}^{d}$, which we write $\mathcal{A}^{co}$ in the following. In particular, under appropriate moment conditions, one can conclude $\rho_{n}(\mathcal{A}^{co}) \to 0$ as $n \to \infty$ if $d^{3/2}/n \to 0$ from Bentkus (2005)’s result. Meanwhile, it has recently attracted much attention in the probabilistic literature to derive bounds for the Wasserstein distances of order $p \geq 1$ between the laws of $S_{n}^{X}$ and $Z_{n}^{X}$ in high-dimensional settings; see Zhai (2018), Bonis (2020), Courtade et al. (2019), Eldan et al. (2020) and Fathi (2019), among others. As illustrated in Zhai (2018, Section 1.1), such a bound can be used to improve the dimension dependence to obtain the convergence $\rho_{n}(\mathcal{A}^{co}) \to 0$ under some situations. For example, when each $X_{j}$ is isotropic and satisfies a Poincaré inequality with constant $C$ independent of $n$, we can deduce $\rho_{n}(\mathcal{A}^{co}) \to 0$ as $n \to \infty$ if $d^{3/2}/n \to 0$ from (Zhai 2018, Proposition 1.4) and (Courtade et al. 2019, Theorem 4.1).

As outlined above, one typically requires sub-linear dependence of $d$ on $n$ to get $\rho_{n}(\mathcal{A}^{co}) \to 0$ or the convergences of the Wasserstein distances. In fact, one can easily verify that this is usually necessary for getting (at least) the latter convergences. Nevertheless, in modern data science, one is often interested in a situation where $d$ is (much) larger than $n$. Recently, the path-breaking work of Chernozhukov et al. (2013) has shown that, if we restrict our attention to the class $\mathcal{A} = \mathcal{A}^{m}$ of sets of the form $A = \{x \in \mathbb{R}^{d} : \max_{j \in J} x_{j} \leq a\}$ for some $a \in \mathbb{R}$ and $J \subset \{1, \ldots, d\}$ ($x_{j}$ denotes the $j$th coordinate of $x$), we can deduce $\rho_{n}(\mathcal{A}^{m}) \to 0$ as $n \to \infty$ under appropriate moment conditions even if $d$ is as large as $e^{C\sqrt{n}}$ for some $c, C > 0$. This type of convergence is indeed enough for many statistical applications in high-dimensional inference such as construction of simultaneous confidence intervals and strong control of the family-wise error rate (FWER) in multiple testing; see Belloni et al. (2018) for details. This result has further been extended in Chernozhukov et al. (2017a) to the case that $\mathcal{A} = \mathcal{A}^{re}$ is the class of all hyperrectangles in $\mathbb{R}^{d}$: $\mathcal{A}^{re}$ consists of all sets $A$ of the form

$$
A = \{x \in \mathbb{R}^{d} : a_{j} \leq x_{j} \leq b_{j} \text{ for all } j = 1, \ldots, d\}
$$

for some $-\infty \leq a_{j} \leq b_{j} \leq \infty$, $j = 1, \ldots, d$. In particular, under suitable moment conditions, they have obtained
\begin{equation}
\rho_n(\mathcal{A}^{re}) \leq C \left( \frac{\log^7(\log)dn}{n} \right)^{1/6}, 
\end{equation}

where \( C > 0 \) is a constant independent of \( n \); see Proposition 2.1 in Chernozhukov et al. (2017a). Indeed, they have also shown that inequality (1) continues to hold true with replacing \( \mathcal{A}^{re} \) by a class of simple convex sets or sparsely convex sets under appropriate assumptions; see Section 3 in Chernozhukov et al. (2017a) for details.

From (1), we infer \( \rho_n(\mathcal{A}^{re}) \to 0 \) as \( n \to \infty \) if \( (\log d)^7/n \to 0 \). Although this condition is much weaker than the ones imposed to obtain the convergence of \( \rho_n(\mathcal{A}^{co}) \) or the Wasserstein distances, it is still unclear whether this condition is necessary to get the convergence \( \rho_n(\mathcal{A}^{re}) \to 0 \) under reasonable moment conditions. In fact, in Chernozhukov et al. (2017a), it is conjectured that \( \log^7 d \) would be replaced by \( \log^3 d \) in (1) (see Remark 2.1 in Chernozhukov et al. (2017a)). In this paper, we show that \( \log^7 d \) can be replaced by \( \log^5 d \) in (1) under the same assumptions as in Chernozhukov et al. (2017a). Moreover, if \( S_n^X \) has a common factor across the components, we can further reduce \( \log^7 d \) in (1) to \( \log^3 d \). Thus, under appropriate moment conditions, we obtain \( \rho_n(\mathcal{A}^{re}) \to 0 \) as \( n \to \infty \) if \( (\log d)^5/n \to 0 \) in a general setting and \( (\log d)^3/n \to 0 \) in the presence of a common factor across the components of \( S_n^X \). Note that it is still unknown whether these conditions are improvable or not in a minimax sense (see the end of Sect. 2 for a discussion).

We shall mention that there are a few relevant studies which intend to relax the dimension growth conditions in convergences related to the above problems: Deng and Zhang (2020) have shown that the condition \( (\log d)^5/n \to 0 \) is sufficient to obtain the consistency of some bootstrap approximations for \( \max_{1 \leq j \leq d} S_n^X_{n,j} \). They have also shown that the Rademacher bootstrap approximation for \( \max_{1 \leq j \leq d} S_n^X_{n,j} \) is consistent if \( (\log d)^4/n \to 0 \) and \( X_i s \) are symmetric. Kuchibhotla et al. (2019) have proved \( \rho_n(\mathcal{A}^{m}) \to 0 \) as \( n \to \infty \) under the condition \( (\log d)^4/n \to 0 \) when the median of \( \max_{1 \leq j \leq d} Z_{n,j}^X \) is tight as \( n \to \infty \). Compared to these existing results, this paper directly improves the dimension growth conditions of some estimates obtained in Chernozhukov et al. (2017a); see Remark 2.1 (see also Remarks 2.2 and 3.2).

The remainder of the paper is organized as follows. Section 2 presents the main results of the paper, while Sect. 3 develops a bootstrap approximation theorem complementing the main results in terms of statistical applications. Section 4 demonstrates a fundamental lemma and its proof. Sections 5 and 6 are devoted to the proofs for the results stated in Sects. 2 and 3.

### 1.1 Notation

Throughout the paper, we assume \( d \geq 3 \) and \( n \geq 3 \). We regard all vectors as column vectors. Given a vector \( x \in \mathbb{R}^d \), we denote by \( x_j \) the \( j \)th coordinate of \( x \), i.e. \( x = (x_1, \ldots, x_d)^T \). Here, \( ^T \) means transposition of a matrix. We write \( \|x\|_{\ell_\infty} = \max_{1 \leq j \leq d} |x_j| \). Given a sequence \( X = (X_i)_{i=1}^n \) of random vectors in \( \mathbb{R}^d \), we denote the \( j \)th component of \( X_i \) by \( X_{ij} \) or \( X_{i,j} \). For a positive integer \( k \), we write \( [k] := \{1, \ldots, k\} \). \( \mathcal{B}(\mathbb{R}) \) denotes the Borel \( \sigma \)-field of \( \mathbb{R} \). For a function \( h : \mathbb{R}^d \to \mathbb{R} \),
we set \( \|h\|_\infty := \sup_{x \in \mathbb{R}^d} |h(x)| \). \( C^m_b(\mathbb{R}^d) \) denotes the space of all \( C^m \) functions all of whose partial derivatives are bounded. We write \( \partial_{x_1 \cdots x_r} = \frac{\partial^r}{\partial x_1 \cdots \partial x_r} \) for short. Given a random variable \( \xi \), we set \( \|\xi\|_p := \{E[|\xi|^p]\}^{1/p} \) for every \( p > 0 \). Also, we define the \( \psi_1 \)-Orlicz norm of \( \xi \) by \( \|\xi\|_{\psi_1} := \inf\{C > 0 : E[\psi_1(|\xi|/C)] \leq 1\} \), where \( \psi_1(x) := \exp(x) - 1 \). For two real numbers \( a \) and \( b \), the notation \( a \lesssim b \) means that \( a \leq cb \) for some universal constant \( c > 0 \).

2 Main results

The following quantities play a key role to deduce our results:

**Definition 2.1** For a random vector \( F \) in \( \mathbb{R}^d \), the concentration function \( C_F : (0, \infty) \to [0, 1] \) is defined by

\[
C_F(\epsilon) := \sup_{y \in \mathbb{R}^d} P\left( 0 \leq \max_{1 \leq j \leq d}(F_j - y_j) \leq \epsilon \right), \quad \epsilon > 0.
\]

We also set

\[
\Theta_X := \sup_{\epsilon > 0} \epsilon^{-1} C_{Z_n^X}(\epsilon).
\]

This definition of the concentration function \( C_F \) is a multivariate extension of the one used in Section 2 of Le Cam (1986, Chapter 15). When \( d = 1 \), \( C_F \) is essentially the same quantity as the Lévy concentration function considered in Chernozhukov et al. (2015, Definition 1). In fact, in this case, we evidently have

\[
C_F(2\epsilon) = \sup_{y \in \mathbb{R}} P(|F - y| \leq \epsilon).
\]

The quantity \( \Theta_X \) measures the degree of anti-concentrations of \( Z_n^X \). As emphasized in Chernozhukov et al. (2013, 2015), it is crucial that \( Z_n^X \) exhibits reasonable anti-concentrations with respect to the dimension \( d \) in order to obtain high-dimensional CLTs.

The following is the main result of the paper.

**Theorem 2.1** Assume \( \Theta_X < \infty \). Assume also \( \max_{1 \leq j \leq d} n^{-1} \sum_{i=1}^n E[X_{ij}^4] \leq B_n^2 \) for some constant \( B_n \geq 1 \). Then the following statements hold true:

(a) If \( \max_{1 \leq i \leq n} \max_{1 \leq j \leq d} \|X_{ij}\|_{\psi_1} \leq B_n \), there is a universal constant \( C > 0 \) such that

\[
\rho_n(A^{re}) \leq C \Theta_X^{2/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}^{1/3} \right),
\]

where

\( \delta_{n,1} \) and \( \delta_{n,2} \) are defined as follows.
\[ \delta_{n,1} := \frac{B_n^2 (\log d)^3}{n}, \quad \delta_{n,2} := \frac{B_n^2 (\log d)^2 (\log n)^2}{n}. \]

(b) If \( \max_{1 \leq i \leq n} \| \max_{1 \leq j \leq d} |X_{ij}| \|_q \leq D_n \) for some \( q \in (2, \infty) \) and \( D_n \geq 1 \), there is a constant \( K_q > 0 \) which depends only on \( q \) such that

\[ \rho_n(A^{\text{rev}}) \leq K_q \Theta_X^{2/3} \left\{ \delta_{n,1}^{1/6} + \delta_{n,2}(q)^{1/3} \right\}, \]

where

\[ \delta_{n,2}(q) := \frac{D_n^2 (\log d)^{2-2/q}}{n^{1-2/q}}. \]

To get meaningful estimates from Theorem 2.1, we need to bound the quantity \( \Theta_X \). The following result, which is called Nazarov’s inequality in Chernozhukov et al. (2017a), can be used for this purpose (see Chernozhukov et al. 2017b for the proof).

Lemma 2.1 (Nazarov’s inequality) Let \( Z \) be a centered Gaussian vector in \( \mathbb{R}^d \) with \( \sigma := \min_{1 \leq j \leq d} \|Z_j\|_2 > 0 \). Then for any \( \epsilon > 0 \),

\[ C_Z(\epsilon) \leq \frac{\epsilon}{\sigma} (\sqrt{2 \log d} + 2). \]

Theorem 2.1 and Lemma 2.1 immediately yield the following result.

Corollary 2.1 Assume \( \sigma := \min_{1 \leq j \leq d} \|S_{n,j}\|_2 > 0 \). Then under the assumptions of Theorem 2.1(a), there is a universal constant \( C > 0 \) such that

\[ \rho_n(A^{\text{rev}}) \leq \frac{C}{\sigma^{2/3}} \left( \frac{B_n^2 (\log d n)^5}{n} \right)^{1/6}. \]

Also, under the assumptions of Theorem 2.1(b), there is a constant \( K_q > 0 \) depending only on \( q \) such that

\[ \rho_n(A^{\text{rev}}) \leq \frac{K_q}{\sigma^{2/3}} \left\{ \left( \frac{B_n^2 (\log d)^5}{n} \right)^{1/6} + \left( \frac{D_n^2 (\log d)^{3-2/q}}{n^{1-2/q}} \right)^{1/3} \right\}. \]

Remark 2.1 Corollary 2.1 improves the bounds given by Chernozhukov et al. (2017a), Proposition 2.1) in terms of dimension dependence under the same assumptions. In particular, we have \( \rho_n(A^{\text{rev}}) \to 0 \) as \( n \to \infty \) if \( (\log d)^5/n = o(1) \), provided that \( \max_{i,j} \|X_{ij}\|_{q_i} = O(1) \) or \( \max_i \| \max_j \|_{4} = O(1) \). As a consequence, we can readily improve the dimension growth conditions in existing results obtained by applications of Proposition 2.1 in Chernozhukov et al. (2017a) (or Corollary 2.1.
in Chernozhukov et al. 2013). For example, the condition $(\log p_1)^7 = o(n)$ imposed in Belloni et al. (2015), Corollary 3 can be replaced by $(\log p_1)^5 = o(n)$. Another example is Condition E in Belloni et al. (2018), Theorem 2.1, where we can replace $\log^7(pn)$ by $\log^5(pn)$.

In some situation, we can bound $\Theta_X$ by a dimension-free constant. This is the case when $S^X_n$ has a common factor across the components:

**Lemma 2.2** Let $Z$ be a centered Gaussian vector in $\mathbb{R}^d$. Also, let $\zeta$ be a standard Gaussian variable independent of $Z$. Let $a_1, \ldots, a_d$ be non-zero real numbers and define $F := (Z_1 + a_1 \zeta, \ldots, Z_d + a_d \zeta)^\top$. Then we have

$$C_F(\varepsilon) \leq \frac{\varepsilon}{\sqrt{2\pi a}}$$

for any $\varepsilon > 0$, where $a := \min_{1 \leq j \leq d} |a_j|$.

Lemma 2.2 is inspired by Lemma 1 in Le Cam (1986), Chapter 15. In fact, if $a_1 = \cdots = a_d$, Lemma 2.2 is obtained as a special case of that lemma.

**Corollary 2.2** Suppose that there is a vector $a \in \mathbb{R}^d$ such that $S^X_n - aa^\top$ is positive semidefinite and $a := \min_{1 \leq j \leq d} |a_j| > 0$. Then under the assumptions of Theorem 2.1(a), there is a universal constant $C > 0$ such that

$$\rho_n(A^{re}) \leq \frac{C}{a^{2/3}} \left( \delta_{n,1}^{1/6} + \delta_{n,2}^{1/3} \right).$$

Also, under the assumptions of Theorem 2.1(b), there is a constant $K_q > 0$ depending only on $q$ such that

$$\rho_n(A^{re}) \leq \frac{K_q}{a^{2/3}} \left( \delta_{n,1}^{1/6} + \delta_{n,2}(q)^{1/3} \right).$$

**Remark 2.2** If we restrict our attention to $\rho_n(A^{re})$, i.e. Gaussian approximation for $\max_{1 \leq j \leq d} S^X_n$, the quantity $\Theta_X$ appearing in Theorem 2.1 can be replaced by the following one:

$$\sup_{\varepsilon > 0} \varepsilon^{-1} \sup_{t \in \mathbb{R}} P \left( 0 \leq \max_{1 \leq j \leq d} Z^X_{n,j} - t \leq \varepsilon \right).$$

To bound this quantity, we can benefit from some recently established anti-concentration inequalities. For example, Theorem 2.2 in Kuchibhotla et al. (2019) develops a dimension-free bound based on the median of $\max_{1 \leq j \leq d} |Z^X_{n,j}|$, while Theorem 3.2 in Belloni and Oliveira (2018) deals with a situation where $\min_{1 \leq j \leq d} ||S^X_{n,j}||_2$ is small (see also the proof of Lopes et al. 2020, Proposition C.1). Such a situation can also be handled by Theorem 10 in Deng and Zhang (2020), which generalizes Lemma 2.1 and can be used to get a potentially better bound for $\Theta_X$; see also Remark 3.2.
By Theorem 2.1, when \( \max_{i,j} \|X_{ij}\|_{\psi_1} = O(1) \) or \( \max_i \max_j \|X_{ij}\|_4 = O(1) \) as \( n \to \infty \), we have \( \rho_n(\mathcal{A}^n) \to 0 \) if \( \Theta_X^2(\log d)^3/n \to 0 \). Then it is interesting to ask whether the condition \( \Theta_X^2(\log d)^3/n \to 0 \) can be weakened or not. Thus, far the answer is not known to the author’s knowledge, but it might be worth mentioning that there is a situation where the condition \( (\log d)^3/n \to 0 \) cannot be weakened:

**Proposition 2.1** Let \( \xi = (\xi_{ij})_{i,j=1}^\infty \) be an array of i.i.d. random variables such that \( \|\xi_{ij}\|_{\psi_1} < \infty \), \( E[\xi_{ij}] = 0 \), \( E[\xi_{ij}^2] = 1 \) and \( \gamma := E[\xi_{ij}^3] < 0 \). Also, let \( \zeta = (\zeta_j)_{j=1}^\infty \) be a sequence of i.i.d. standard normal variables. Then if the sequence \( d_n \in \mathbb{N} \) satisfies \( (\log d_n)^3/n \to c < 0 \) as \( n \to \infty \), we have

\[
\limsup_{n \to \infty} \sup_{x \in \mathbb{R}} \left| P\left( \max_{1 \leq i \leq d_n} \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi_{ij} \leq x \right) - P\left( \max_{1 \leq i \leq d_n} \zeta_j \leq x \right) \right| > 0.
\]

The proof of Proposition 2.1 is based on a Cramér type large deviation result, and it has already been mentioned (at least informally) in the literature; see, e.g. Hall (2006) (see also Remark 1 in Chen 2018). Note that the proposition does not imply that the condition \( \Theta_X^2(\log d)^3/n \to 0 \) is necessary because \( \Theta_X \) is of order \( \sqrt{\log d} \) under the assumptions of the proposition; see Example 2 in Chernozhukov et al. (2015).

### 2.1 Comparison to Kuchibhotla et al. (2019)’s results

Kuchibhotla et al. (2019) have established Gaussian approximation for \( \max_{1 \leq i \leq d} |\mathcal{X}_{n,j}| \) under a variety of assumptions. In particular, their Theorem 3.2 reads as follows: Assume \( X_1, \ldots, X_n \) are i.i.d. and have unit variances for simplicity. Then for any \( q \geq 3 \),

\[
\sup_{x \in \mathbb{R}} \left| P\left( \max_{1 \leq i \leq d} |\mathcal{X}_{n,j}| \leq x \right) - P\left( \max_{1 \leq i \leq d} |Z_{n,j}| \leq x \right) \right| \leq C \left( \frac{1}{2n} + \mu \left( \frac{L_n^2 \log^4 d}{n} \right)^{1/6} + \nu_q \left( \frac{(\log d)^{1-1/q}}{n^{1/2-1/q}} \right) \right).
\]

Here, \( C \) is a universal constant and \( \mu \) is the median of \( \max_{1 \leq i \leq d} |Z_{n,j}| \), while \( L_n \) and \( \nu_q \) are defined as follows: for every \( i = 1, \ldots, n \), let \( \zeta_i \) be the signed Borel measure on \( \mathbb{R}^d \) defined by \( \zeta_i(A) = P(X_i \in A) - P(Z_n^X \in A) \) for every Borel set \( A \) in \( \mathbb{R}^d \). Then we set

\[
L_n := \frac{1}{n} \sum_{i=1}^n \max_{1 \leq i \leq d} \int_{\mathbb{R}^d} |x_j|^3 |\zeta_i|(dx), \quad \nu_q := \left\{ \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^d} \max_{1 \leq i \leq d} |x_j|^q |\zeta_i|(dx) \right\}^{1/q}
\]
where $|\xi|$ denotes the total variation measure of $\xi$. To make comparison to our result possible, we consider the worst case $\mu \gtrsim \sqrt{\log d}$ as in Corollary 2.1 in the following. Then under the assumptions of Theorem 2.1(b), $v_q$ can be bounded as follows:

$$v_q \leq \frac{1}{n} \sum_{i=1}^{n} E[\|X_i\|_{\ell_\infty}^q + \|Z^X_n\|_{\ell_\infty}^q] \leq D_n^q + E[\|Z^X_n\|_{\ell_\infty}^q].$$

Thus, up to a constant depending only on $q$, (2) can be reduced to the following form:

$$\left(\frac{L_n^2 \log^7 d}{n}\right)^{1/6} + \frac{D_n (\log d)^{3/2 - 1/q}}{n^{1/2 - 1/q}} + \frac{E[\|Z^X_n\|_{\ell_\infty}^3]^{1/2} (\log d)^{3/2 - 1/q}}{n^{1/2 - 1/q}}. \quad (3)$$

On the one hand, this bound always exhibits a better dependence on $n$ than our result because the latter contains a bound of the form

$$\frac{D_n (\log d)^{3/2 - 1/q}}{n^{1/2 - 1/q}}.$$ 

On the other hand, to ensure that the bounds converge to 0 as $n \to \infty$, our result imposes conditions on $d$ and $D_n$ no worse than (2) as long as both $L_n$ and $B_n$ do not increase with $n$. In particular, when $q \geq 4$ and $D_n$ does not increase with $n$ as well, our bound converges to 0 as $n \to \infty$ if $\log d = o(n^{1/5})$, requiring a weaker condition than the one to ensure convergence of (2).

One clear advantage of Kuchibhotla et al. (2019)'s results is that they are applicable to the situation where only the third moments of $X_i$ are finite. Indeed, their results can handle the case that only the $(2 + \tau)$th moments of $X_i$ are finite for some $\tau > 0$; see discussions after their Theorem 3.1. Meanwhile, we remark that the difference between $L_n$ and $B_n$ could be significant in some statistical applications. For example, let us consider the following nonparametric regression with regular design:

$$y_i = f(i/n) + \epsilon_i, \quad i = 1, \ldots, n,$$

where $f : [0, 1] \to \mathbb{R}$ and $\epsilon_1, \ldots, \epsilon_n$ are centered i.i.d. variables. Let us estimate the function $f$ by the following kernel estimator:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} y_i K\left(\frac{i/n - x}{h_n}\right), \quad x \in [0, 1],$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(u)du = 1$ and $h_n$ is a bandwidth parameter which tends to 0 as $n \to \infty$. For simplicity, we assume $K$ is continuous and compactly supported. To construct uniform confidence bands for $f$ at the points $x_1, \ldots, x_d \in [0, 1]$, we wish to establish Gaussian approximation for the following quantity:

$$\max_{1 \leq j \leq d} \sqrt{n h_n} \left| \hat{f}_n(x_j) - E[\hat{f}_n(x_j)] \right|.$$
For this purpose, we can apply a high-dimensional CLT for $S_X^n$ with $X_{ij} = h^{-1/2} \epsilon_i K((i/n - x_j)/h_n)$. In this case, $L_n$ is typically of order $h_n^{-3/2}$, while

$$\frac{1}{n} \sum_{i=1}^n E[X_{ij}^4] \approx E[\epsilon_1^4] \cdot h_n^{-2} \int_0^1 K\left(\frac{y-x_j}{h_n}\right)^4 dy \leq E[\epsilon_1^4] \cdot h_n^{-1} \int K(u)^4 du,$$

so $B_n$ is of order $h_n^{-1/2}$ as long as $E[\epsilon_4] < \infty$. Note that $D_n$ is also of order $h_n^{-1/2}$ in this situation. See also Remark 4.8 of Koike (2019a) for another application of this type.

We should also mention that Kuchibhotla et al. (2019) also deal with the case that $X_{ij}$ are sub-Weibull for all $i, j$, which generalizes the assumptions of Theorem 2.1(a); see their Corollary 3.1 for details. Moreover, they have also developed high-dimensional versions of non-uniform CLTs and Cramér type large deviations. These topics are beyond the scope of this paper.

### 3 Bootstrap approximation

In terms of statistical applications, the Gaussian approximation results obtained in the previous section are infeasible unless the covariance matrix $\mathbb{C}_X^n$ is known for statisticians. Moreover, even if this is the case, the probability $P(Z_X^n \in A)$ is analytically intractable for a general set $A \in \mathcal{A}^c$. For these reasons, this section develops bootstrap approximation for $P(Z_X^n \in A)$ with $A \in \mathcal{A}^c$, following Chernozhukov et al. (2017a).

Let $w = (w_i)_{i=1}^n$ be a sequence of independent random variables independent of $X$. We consider the wild bootstrap (also called the multiplier bootstrap) with multiplier variables $w$ as

$$S_n^{WB} = (S_{n,1}^{WB}, \ldots, S_{n,d}^{WB})^\top := \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i (X_i - \bar{X}),$$

where $\bar{X} := n^{-1} \sum_{i=1}^n X_i$. We set

$$\rho_n^{WB}(A^c) := \sup_{A \in \mathcal{A}^c} \left| P(S_n^{WB} \in A \mid X) - P(Z_X^n \in A) \right|.$$

**Theorem 3.1** Suppose that $E[w_i] = 0$ and $E[w_i^2] = 1$ for every $i$. Suppose also that there is a constant $b \geq 1$ such that $|w_i| \leq b$ a.s. for every $i$. Then the following statements hold true:

(a) Under the assumptions of Theorem 2.1(a), there is a universal constant $C' > 0$ such that

$$E[\rho_n^{WB}(A^c)] \leq C' \Theta_X^{2/3} \left( (b^2 \delta_{n,1})^{1/6} + (b^2 \delta_{n,2})^{1/3} \right).$$
(b) Under the assumptions of Theorem 2.1(b), there is a constant $K'_q > 0$ depending only on $q$ such that
\[
E[\rho_n^{\text{WB}}(A^\text{re})] \leq K'_q \Theta^{2/3}_X \left( (b^2 \delta_{n,1})^{1/6} + (b^2 \delta_{n,2}(q))^{1/3} \right).
\]

Theorem 3.1 and Lemma 2.1 yield the following counterpart of Corollary 2.1.

**Corollary 3.1** Assume $\sigma := \min_{1 \leq j \leq d} \| S_{n,j}^X \|_2 > 0$. Then under the assumptions of Theorem 3.1(a), there is a universal constant $C' > 0$ such that
\[
E[\rho_n^{\text{WB}}(A^\text{re})] \leq \frac{C'}{\sigma^{2/3}} \left( \frac{b^2 B_n^2 (\log dn)^5}{n} \right)^{1/6}.
\] (4)

Also, under the assumptions of Theorem 3.1(b), there is a constant $K'_q > 0$ depending only on $q$ such that
\[
E[\rho_n^{\text{WB}}(A^\text{re})] \leq \frac{K'_q}{\sigma^{2/3}} \left\{ \left( \frac{b^2 B_n^2 (\log dn)^5}{n} \right)^{1/6} + \left( \frac{b^2 D_n^2 (\log dn)^3 - 2/q}{n^{1-2/q}} \right)^{1/3} \right\}. \quad (5)
\]

It is of course possible to derive a bootstrap counterpart of Corollary 2.2. We omit the precise statement.

**Remark 3.1** (Relation to Chernozhukov et al. 2017a) Chernozhukov et al. (2017a) have established similar results to Corollary 3.1 when $w$ is Gaussian. Indeed, they have derived stronger results that the probabilities of $\rho_n^{\text{WB}}(A^\text{re}(d))$ exceeding the right hand sides of (4) or (5) are small; see Proposition 4.1 in Chernozhukov et al. (2017a) for details. It is presumably possible to obtain similar results in our case by showing that the variables appearing in the right side of (32) concentrate at their expectations.

**Remark 3.2** (Relation to Deng and Zhang 2020) Deng and Zhang (2020) have developed analogous results to Corollary 3.1 for instead of $\rho_n^{\text{WB}}(A^\text{re})$ and assuming $w_j$ are i.i.d. and sub-Gaussian (rather than bounded) but satisfies $E[w_j^2] = 1$. In particular, under the assumptions of Theorem 2.1(a), we can derive the following bound from their Corollary 1(i):
\[
E \left[ \sup_{A \in A^n} \left| P \left( S_n^{\text{WB}} \in A \mid X \right) - P \left( S_n^X \in A \right) \right| \right] \leq \frac{C^*}{\sigma^{2/3}} \left( \frac{B_n^2 (\log dn)^5}{n} \right)^{1/6}, \quad (6)
\]
where $C^*$ depends only on the sub-Gaussian parameter of $w_1$ and

$$
\bar{\sigma} := \min_{1 \leq j \leq d} \frac{2 + \sqrt{2 \log d}}{1/\sigma_{(1)} + (1 + \sqrt{2 \log j})/\sigma_{(j)}}$

with $\sigma_{(1)} \leq \cdots \leq \sigma_{(d)}$ being the ordered diagonal entries of $\mathbb{G}^X_n$. We can also derive the following bound from their Corollary 3(ii) under the assumptions of Theorem 2.1(b) with taking $\delta = \left( \frac{B_n^2 (\log d n)^{5/6}}{\sigma_n^4} \right)^{1/6}$ in their result (note that $B_n^2 / \sigma^4 \geq 1$ by Jensen’s inequality):

$$
E \left[ \sup_{A \in A_n} \left| P(S_n^{WB} \in A \mid X) - P(S_n^X \in A) \right| \right] \leq \frac{K_q^*}{\bar{\sigma}^2/3} \left( \frac{B_n^2 (\log d n)^{5/6}}{n} \right)^{1/6},
$$

where $K_q^*$ depends only on $q$ and the sub-Gaussian parameter of $w_1$. Since Theorem 10 in Deng and Zhang (2020) implies $\Theta_X \asymp \bar{\sigma}^{-1} \sqrt{\log d}$, our Theorems 2.1(a) and 3.1(a) lead to a bound similar to (6). In the meantime, the bound obtained from our Theorems 2.1(b) and 3.1(b) is worse than (7) due to the presence of the second term. The advantage of our results over Deng and Zhang (2020)’s ones is that we do not need to assume $E[w_{ij}^3] = 1$. We note that they have indeed established stronger estimates as the one stated in Remark 3.1 and allow $w_i$ to be sub-Gaussian (rather than bounded). In this paper, we do not pursue such generalization for simplicity.

**Remark 3.3 (Empirical bootstrap)** It seems difficult to derive a result comparable to Theorem 3.1 for Efron’s empirical bootstrap using our proof technique. This is because we need to bound a quantity of the form $E[\sum_{i=1}^n \max_{1 \leq j \leq d} X_{ij}]$ rather than $E[\max_{1 \leq j \leq d} \sum_{i=1}^n X_{ij}]$ while we apply our key Lemma 4.1 in order to derive such a result for Efron’s empirical bootstrap. We shall remark that this issue has also been pointed out in Deng and Zhang (2020) (see discussions after Theorem 2 of the paper).

### 4 Fundamental lemma

The basic strategy for the proofs of the main results is the same as the one used in the proof of Chernozhukov et al. (2013, Theorem 2.2), which is based on Chernozhukov et al. (2013, Theorem 2.1) and an anti-concentration inequality. Here, since we do not explicitly bound the quantity $\Theta_X$, we need to establish only a counterpart of the former. This part is the main technical development of this paper and the result is given as follows:

**Lemma 4.1** Let $Z$ be a centered Gaussian vector in $\mathbb{R}^d$ with covariance matrix $\mathbb{G} = (\mathbb{G}_{jk})_{1 \leq j, k \leq d}$. Suppose that $E[X_{ij}^4] < \infty$ for all $i, j$. Then there is a universal constant $C > 0$ such that
\[ P\left( \max_{1 \leq j \leq d} (S_{n,j}^X - y_j) \in A \right) \leq P\left( \max_{1 \leq j \leq d} (Z_j - y_j) \in A^\epsilon \right) + C \left\{ \epsilon^{-2} \left( \Delta_{n,0}^X \log d + \Delta_{n,1}^X \sqrt{3(\log d)^3 \over n} \right) + \epsilon^{-4} \Delta_{n,2}^X (\log d)^3 \right\} \] (8)

for any \( y \in \mathbb{R}^d, \epsilon > 0 \) and \( A \in \mathcal{B}(\mathbb{R}) \), where

\[ \Delta_{n,0}^X := \max_{1 \leq i, k \leq d} \left| \frac{1}{n} \sum_{j=1}^n E[X_{ij}X_{ik}] - C_{jk} \right|, \quad \Delta_{n,1}^X := \sqrt{\frac{1}{n} E \left[ \max_{1 \leq i \leq d} \sum_{j=1}^n X_{ij}^4 \right]} \]

and

\[ \Delta_{n,2}^X (\epsilon) := \frac{1}{n} \sum_{i=1}^n E \left[ \|X_i\|_{\ell^4} \|X_i\|_{\ell^\infty} > \sqrt{n\epsilon / (3 \log d)} \right]. \]

Lemma 4.1 can be seen as a variant of Chernozhukov et al. (2014, Theorem 4.1) and Chernozhukov et al. (2016, Theorem 3.1), and it is closely related to Gaussian couplings for \( \max_{1 \leq j \leq d} (S_{n,j}^X - y_j) \); see Lemma 4.1 in Chernozhukov et al. (2014). The proof strategy is basically the same as these two theorems and consists of the following two steps: First, we approximate the indicator function \( 1_A \) and the maximum function by appropriate smooth functions. Second, we estimate \( |E[g(S_n^X - y)] - E[g(Z - y)]| \) for a particular class of smooth functions \( g \) and establish their “good” bounds with respect to \( d \). To get good bounds in the second step, we partially follow the idea of Deng and Zhang (2020), where a randomized version of the Lindeberg method is developed to improve the dimension dependence of bootstrap approximations for \( \max_{1 \leq j \leq d} S_{n,j}^X \). To transfer this improvement to Gaussian approximations for \( \max_{1 \leq j \leq d} S_{n,j}^X \), we show that the dimension dependence of Gaussian approximations is improvable for a specific wild bootstrap version of \( S_n^X \) by the Stein kernel method. All together, we will complete the proof of Lemma 4.1.

Indeed, the main contribution of this paper is the final part in the above. To be precise, we can prove the following result:

**Proposition 4.1** Let \( (\eta_i)_{i=1}^n \) be a sequence of i.i.d. random variables independent of \( X \) such that the law of \( \eta_i \) is the beta distribution with parameters 1/2, 3/2 for every \( i \). That is, \( \eta_i \) has the density function of the form \( f(x) = B(1/2, 3/2)^{-1} \sqrt{(1-x)/x}^1 (1, 3/2) \), where \( B(u, v) \) denotes the beta function. Set \( \xi_i := 4\eta_i - 1 \) and \( Y_i := \xi_i X_i \) for every \( i = 1, \ldots, n \) and let \( Y := (Y_i)_{i=1}^n \). Then under the assumptions of Lemma 4.1,

\[ \rho_{h,\beta}(S_n^X, Z) \leq \frac{3}{2} \left( \max_{1 \leq i \leq 2} \beta^{2-i} \|h^{(i)}\|_\infty \right) \left\{ \Delta_{n,0}^X + 5 \sqrt{2 \log(2d^2) \over n} \Delta_{n,1}^X \right\} \]

for any \( h \in C_b^\infty (\mathbb{R}) \) and \( \beta > 0 \), where \( \rho_{h,\beta}(S_n^X, Z) \) is defined by (11).
Proposition 4.1 is proved at the end of Sect. 4.3. Roughly speaking, this result means that a special wild bootstrap version of $S_n^X$ may enjoy Gaussian approximation with an improved dimension dependence condition. Here, the key fact is that the multiplier sequence $\xi_i$ in the above satisfies $E[\xi_i] = 0$ and $E[\xi_i^2] = E[\xi_i^3] = 1$. Hence, we can consider a Lindeberg type interpolation between $S_n^X$ and $S_n^X$ up to the third moments rather than the second moments. This is the main difference between our method and the Stein–Slepian approach by Chernozhukov et al. (2017a) because the latter allows moment matching only up to the second moments.

In fact, if $\| \max_{1 \leq j \leq d} |X_{ij}|_4 \|$ is bounded by a constant $B_n$ uniformly in $i$, we can use a standard Lindeberg method to obtain the following result, which is weaker than Lemma 4.1 but still leads to the dimension improvement in some cases.

**Proposition 4.2** Suppose that there is a constant $B_n$ satisfying $\| \max_{1 \leq j \leq d} |X_{ij}|_4 \| \leq B_n$ for all $i = 1, \ldots, n$. Under the assumptions of Lemma 4.1,

$$P \left( \max_{1 \leq j \leq d} (S_n^X - y_j) \in A \right) \leq P \left( \max_{1 \leq j \leq d} (Z_j - y_j) \in A \right) + C \epsilon^{-2} \left( \Delta_n^X \log d + \sqrt{\frac{B_n^4 (\log d)^3}{n}} \right)$$

for any $y \in \mathbb{R}^d$, $\epsilon > 0$ and $A \in \mathcal{B}(\mathbb{R})$.

We can indeed use this result to improve the dimension dependence of Gaussian approximation for $S_n^X$:

**Corollary 4.1** Assume $\Theta_X < \infty$. Under the assumptions of Proposition 4.2, we have

$$\rho_n(A^{ce}) \leq C \Theta_X^{2/3} \left( \frac{B_n^4 (\log d)^3}{n} \right)^{1/6},$$

where $C > 0$ is a universal constant. In particular, if $\sigma := \min_{1 \leq j \leq d} \| S_n^X \|_2 > 0$, there is a universal constant $C' > 0$ such that

$$\rho_n(A^{ce}) \leq \frac{C'}{\sigma^{2/3}} \left( \frac{B_n^4 (\log d)^5}{n} \right)^{1/6}.$$
which is bounded by $B_n^4/n$; see Lemma A.1. Indeed, utilizing the stability property of the smooth max function given by Lemma 4.3 (iii), we can reduce the above bound to the following one (plus some reminder terms):

$$\frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \left[ \max_{1 \leq j \leq d} X_{ij}^4 \right],$$

This is a standard argument in the Chernozhukov–Chetverikov–Kato theory. Now, in this expression, the sum "\[\sum_{i=1}^n\]" comes from the Lindeberg interpolation with replacing $X_1, \ldots, X_n$ by $Y_1, \ldots, Y_n$ (defined in Proposition 4.1) one by one with this order. However, there is no reason to keep the order of replacement because $\sum_{i=1}^n X_i$ is invariant under permutation of $X_1, \ldots, X_n$. This intuitively suggests that we could change the bound (9) to $n^{-1} \max_{1 \leq j \leq d} \mathbb{E}[X_{ij}^4]$ by randomizing the order of replacement, where $I$ is a uniform variable over $\{1, \ldots, n\}$ independent of $X$. This quantity is bounded by $B_n^2/n$ because

$$\frac{1}{n} \max_{1 \leq j \leq d} \mathbb{E}[X_{ij}^4] = \frac{1}{n^2} \max_{1 \leq j \leq d} \sum_{i=1}^n \mathbb{E}[X_{ij}^4].$$

**Remark 4.1** (Application to empirical processes) As developed in Chernozhukov et al. (2014, 2016), it will be possible to apply Lemma 4.1 for obtaining Gaussian approximations for suprema of empirical processes. We remark that this could improve the convergence rate of such an approximation since the term multiplied by $\varepsilon^{-4}$ is often dominated by the term multiplied by $\varepsilon^{-2}$ in (8) under suitable moment conditions as in Lemmas 5.7 and 5.8; see Remark 4.8 in Koike (2019a) for an explanation of why this improves the convergence rate. Nevertheless, this topic is beyond the scope of this paper and left for future work.

The remainder of this section is devoted to the proof of Lemma 4.1.

### 4.1 Smooth approximation

We begin by approximating the indicator function $1_A$ and the maximum function by smooth functions. For the indicator function, we will use the following result.

**Lemma 4.2** (Chernozhukov et al. 2016, Lemma 5.1) For any $\varepsilon > 0$ and Borel set $A$ of $\mathbb{R}$, there is a $C^\infty$ function $h : \mathbb{R} \to \mathbb{R}$ satisfying the following conditions:

(i) There is a universal constant $C > 0$ such that $\|h^{(r)}\|_\infty \leq C \varepsilon^{-r}$ for $r = 1, 2, 3, 4$.

(ii) $1_A(x) \leq h(x) \leq 1_{A^\varepsilon}(x)$ for all $x \in \mathbb{R}$.
Remark 4.2 Formally, Lemma 5.1 in Chernozhukov et al. (2016) states that condition (i) in the above is satisfied only for \( r = 1, 2, 3 \), but the function constructed there indeed satisfies this condition for \( r = 4 \).

Next we introduce the following special form of smooth approximation of the maximum function: for each \( \beta > 0 \), we define the function \( \Phi_\beta : \mathbb{R}^d \rightarrow \mathbb{R} \) by

\[
\Phi_\beta(x) = \beta^{-1} \log \left( \sum_{j=1}^{d} e^{\beta x_j} \right).
\]

This “smooth max function” is one of the key constituents in the Chernozhukov–Chetverikov–Kato theory. One can easily verify the following inequality (cf. Eq.(1) in Chernozhukov et al. (2015)):

\[
0 \leq \Phi_\beta(x) - \max_{1 \leq j \leq d} x_j \leq \beta^{-1} \log d \tag{10}
\]

for any \( x \in \mathbb{R}^d \). Thus, \( \Phi_\beta \) better approximates the maximum function as the value of \( \beta \) increases. The next lemma summarizes Lemmas 5–6 in Deng and Zhang (2020) and highlights the key properties of this smooth max function:

Lemma 4.3 For any \( \beta > 0 \), \( m \in \mathbb{N} \) and \( C^m \) function \( h : \mathbb{R} \rightarrow \mathbb{R} \), there is an \( \mathbb{R}^m \)-valued function \( \Upsilon_\beta(x) = (\Upsilon_{j_1}^{j_m}(x))_{1 \leq j_1, \ldots, j_m \leq d} \) on \( \mathbb{R}^d \) satisfying the following conditions:

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

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

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

where \( c_m > 0 \) depends only on \( m \).

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

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

As a result, given two random variables \( F \) and \( G \), we explore bounds for the quantity

\[
\rho_{h, \beta}(F, G) := \sup_{y \in \mathbb{R}^d} \left| \mathbb{E}[h(\Phi_\beta(F-y))] - \mathbb{E}[h(\Phi_\beta(G-y))] \right| \tag{11}
\]

for a (smooth) bounded function \( h : \mathbb{R} \rightarrow \mathbb{R} \) and \( \beta > 0 \) in the following.
4.2 Randomized Lindeberg method

The next lemma essentially has the same content as (Deng and Zhang 2020, Theorem 5). For the sake of completeness, we give a self-contained proof.

In the following, we will use the standard multi-index notation: For a multi-index \( \lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{Z}_+^d \), we set \( |\lambda| := \lambda_1 + \cdots + \lambda_d \). \( \lambda! := \lambda_1! \cdots \lambda_d! \) and \( \partial^\lambda := \partial_{\lambda_1} \cdots \partial_{\lambda_d} \) as usual. Also, given a vector \( x = (x_1, \ldots, x_d) \in \mathbb{R}^d \), we write \( x^\lambda = x_1^{\lambda_1} \cdots x_d^{\lambda_d} \).

**Lemma 4.4** Let \( X = (X_i)_{i=1}^n \) and \( Y = (Y_i)_{i=1}^n \) be two sequences of independent centered random vectors in \( \mathbb{R}^d \). Suppose that there is an integer \( m \geq 3 \) such that
\[
E[|X_{ij}|^{m} + |Y_{ij}|^{m}] < \infty \quad \text{for all} \quad i \in [N], \quad j \in [d] \quad \text{and} \quad E[X_i^\lambda] = E[Y_i^\lambda] \quad \text{for all} \quad i \in [N] \quad \text{and} \quad \lambda \in \mathbb{Z}_+^d \quad \text{with} \quad |\lambda| \leq m - 1.
\]
Then for any \( h \in C_b^m(\mathbb{R}) \) and \( \beta > 0 \), we have
\[
\rho_{h, \beta}(S_n^X, S_n^Y)
\leq C m n^{-\frac{m}{2}} \left( \max_{1 \leq i \leq m} \|h^{(i)}\|_\infty \right) \left\{ \max_{1 \leq i \leq d} \sum_{i=1}^n E[|X_{ij}|^{m} + |Y_{ij}|^{m}] + \sum_{i=1}^n E[\|X_i\|^{m} \vee \|Y_i\|^{m}] > \sqrt{n}/\beta \right\} \max_{1 \leq i \leq d} E[|X_{ij}|^{m} + |Y_{ij}|^{m}]
\]
where \( C_m > 0 \) depends only on \( m \).

**Proof** We may assume that \( X \) and \( Y \) are independent without loss of generality. Throughout the proof, for two real numbers \( a \) and \( b \), the notation \( a \lesssim_m b \) means that \( a \leq c_m b \) for some constant \( c_m > 0 \) which depends only on \( m \).

Take a vector \( y \in \mathbb{R}^d \) and define the function \( \Psi : \mathbb{R}^d \to \mathbb{R} \) by \( \Psi(x) = h(\Phi(x - y)) \) for \( x \in \mathbb{R}^d \). We denote by \( \mathcal{G}_n \) the set of all permutations of \([n]\). For any \( \sigma \in \mathcal{G}_n \) and \( k \in [n] \), we set
\[
S_n^\sigma(k) := \frac{1}{\sqrt{n}} \sum_{i=1}^k X_{\sigma(i)} + \frac{1}{\sqrt{n}} \sum_{i=k+1}^n Y_{\sigma(i)} \quad \text{and} \quad \tilde{S}_n^\sigma(k) := S_n^\sigma(k) - \frac{1}{\sqrt{n}} X_{\sigma(k)}.
\]
By construction \( \tilde{S}_n^\sigma(k) \) is independent of \( X_{\sigma(k)} \) and \( Y_{\sigma(k)} \). We also have
\[
S_n^\sigma(k) = \tilde{S}_n^\sigma(k) + n^{-1/2} X_{\sigma(k)} \quad \text{and} \quad S_n^\sigma(k - 1) = \tilde{S}_n^\sigma(k) + n^{-1/2} Y_{\sigma(k)} \quad \text{(with} \quad S_n^\sigma(0) := n^{-1/2} \sum_{i=1}^n Y_{\sigma(i)}).\]
Moreover, it holds that \( S_n^\sigma(n) = S_n^X \) and \( S_n^\sigma(0) = S_n^Y \). Therefore, we have

(Springer)
Lemma 4.3 yields

\[
\left| E \left[ \Psi(S_n^X) \right] - E \left[ \Psi(S_n^Y) \right] \right| = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \left| E \left[ \Psi(S_n^\sigma(0)) \right] - E \left[ \Psi(S_n^\sigma(1)) \right] \right|
\]

\[
\leq \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \sum_{k=1}^{n} \left| E \left[ \Psi(S_n^\sigma(k)) \right] - E \left[ \Psi(S_n^\sigma(k-1)) \right] \right|.
\]

(12)

Now, when \( W = X \) or \( W = Y \), Taylor’s theorem and the independence of \( W_{\sigma(k)} \) from \( \hat{S}_n^\sigma(k) \) yield

\[
E \left[ \Psi(\hat{S}_n^\sigma(k) + n^{-1/2} W_{\sigma(k)}) \right] = \sum_{\lambda \in \mathbb{Z}^d_+ : |\lambda| \leq m-1} \frac{n^{-|\lambda|/2}}{\lambda!} E \left[ \partial^{\lambda} \Psi(\hat{S}_n^\sigma(k)) \right] E \left[ W_{\sigma(k)}^{\lambda} \right] + R_k^\sigma [W],
\]

where

\[
R_k^\sigma [W] := n^{-m/2} \sum_{\lambda \in \mathbb{Z}^d_+ : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1} E \left[ \partial^{\lambda} \Psi(\hat{S}_n^\sigma(k) + m^{-1/2} W_{\sigma(k)}) W_{\sigma(k)}^{\lambda} \right] dt.
\]

Since we have \( E[X_i^j] = E[Y_i^j] \) for all \( i \in [N] \) and \( \lambda \in \mathbb{Z}^d_+ \) with \( |\lambda| \leq m-1 \) by assumption, we obtain

\[
\left| E \left[ \Psi(S_n^\sigma(k)) \right] - E \left[ \Psi(S_n^\sigma(k-1)) \right] \right| \leq |R_k^\sigma [X]| + |R_k^\sigma [Y]|.
\]

(13)

We estimate \( R_k^\sigma [W] \) as

\[
|R_k^\sigma [W]| \leq I_k^\sigma [W] + II_k^\sigma [W],
\]

where

\[
I_k^\sigma [W] := n^{-m/2} \sum_{\lambda \in \mathbb{Z}^d_+ : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1}
E \left[ \left| \partial^{\lambda} \Psi(\hat{S}_n^\sigma(k) + m^{-1/2} W_{\sigma(k)}) W_{\sigma(k)}^{\lambda} \right| \right] dt,
\]

\[
II_k^\sigma [W] := n^{-m/2} \sum_{\lambda \in \mathbb{Z}^d_+ : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1-t)^{m-1}
E \left[ \left| \partial^{\lambda} \Psi(\hat{S}_n^\sigma(k) + m^{-1/2} W_{\sigma(k)}) W_{\sigma(k)}^{\lambda} \right| \right] dt.
\]

(14)

Lemma 4.3 yields
Now we focus on \( I_k^c [W] \). We rewrite it as

\[
I_k^c [W] = \frac{n^{-m}}{(m-1)!} \sum_{j_1, \ldots, j_m=1}^{d} \int_0^1 (1-t)^{m-1} \times E \left[ \partial_{j_1 \ldots j_m} \Psi \left( \hat{S}_n^a (k) + m^{-1/2} W_{\sigma(k)} \right) \prod_{l=1}^{m} W_{\sigma(k),j_l} \right] \| W_{\sigma(k)} \|_{\ell_{\infty}} \leq \sqrt{n/\beta} \right] dr.
\]

We have by the AM–GM inequality

\[
\left| \prod_{l=1}^{m} W_{\sigma(k),j_l} \right| \leq \frac{1}{m} \sum_{l=1}^{m} |W_{\sigma(k),j_l}|^m.
\]

Inserting this inequality into the first formula and noting that \( \partial_{j_1 \ldots j_m} \Psi \) does not depend on the order of \( j_1, \ldots, j_m \), we obtain

\[
I_k^c [W] \leq \frac{n^{-m}}{(m-1)!} \sum_{j_1, \ldots, j_m=1}^{d} \int_0^1 (1-t)^{m-1} \times E \left[ \partial_{j_1 \ldots j_m} \Psi \left( \hat{S}_n^a (k) + m^{-1/2} W_{\sigma(k)} \right) \right] \left[ \left( W_{\sigma(k),j_l} \right)^m \right] \| W_{\sigma(k)} \|_{\ell_{\infty}} \leq \sqrt{n/\beta} \right] dr.
\]

Thus, by Lemma 4.3, we obtain

\[
I_k^c [W] \leq n^{-m} \sum_{j_1, \ldots, j_m=1}^{d} E \left[ \left| Y_{j_1 \ldots j_m} \left( \hat{S}_n^a (k) - y \right) \right| W_{\sigma(k),j_l} \right] \right|^m
\]

\[
= n^{-m} \sum_{j_1, \ldots, j_m=1}^{d} E \left[ \left| Y_{j_1 \ldots j_m} \left( \hat{S}_n^a (k) - y \right) \right| \right] E[|W_{\sigma(k),j_l}|^m]
\]

\[
= n^{-m} \sum_{j_1, \ldots, j_m=1}^{d} \left\{ E \left[ \left| Y_{j_1 \ldots j_m} \left( \hat{S}_n^a (k) - y \right) \right| ; \mathcal{E}_{\sigma,k} \right] + E \left[ \left| Y_{j_1 \ldots j_m} \left( \hat{S}_n^a (k) - y \right) \right| ; \mathcal{E}_{\sigma,k}^c \right] \right\} E[|W_{\sigma(k),j_l}|^m]
\]

\[
= A_k^c [W] + B_k^c [W],
\]

where \( \mathcal{E}_{\sigma,k} := \{ \| X_{\sigma(k)} \|_{\ell_{\infty}} \lor \| Y_{\sigma(k)} \|_{\ell_{\infty}} \leq \sqrt{n/\beta} \} \). To estimate \( A_k^c [W] \), we adopt an argument analogous to the proof of Eq. (6.7) in Koike (2019b), which is inspired by the proof of (Deng and Zhang 2020, Lemma 2). Let \( (\delta_l)_{l=1}^{\infty} \) be a sequence of i.i.d. Bernoulli variables independent of \( X \) and \( Y \) with \( P(\delta_l = 1) = 1 - P(\delta_l = 0) = i/(n+1) \). We set \( \zeta_{k,i} := \delta_k X_i + (1 - \delta_k) Y_i \) for all \( k, i \in [n] \). Then Lemma 4.3 yields
Next, for any $k, i \in [n]$, we set
\[ A_{k,i} = \{(A, B) : A \subset [n], B \subset [n], A \cup B = [n] \setminus \{i\}, \#A = k - 1, \#B = n - k\}, \]
where $\#S$ denotes the number of elements in a set $S$. We also set
\[ A_k = \{(A, B) : A \subset [n], B \subset [n], A \cup B = [n], \#A = k, \#B = n - k\} \]
for every $k \in \{0, 1 \ldots, n\}$. Moreover, for any $(A, B) \in \bigcup_{k=0}^{n} A_k$ and $k \in [n]$, we set
\[ W_{k}^{(A,B)} = \begin{cases} X_k & \text{if } k \in A, \\ Y_k & \text{if } k \in B. \end{cases} \]
Then we define $S_n^{(A,B)} := \sum_{k=1}^{n} W_{k}^{(A,B)}$ and $\widehat{S}_n^{(A,B)}(i) := \sum_{k=1,k\neq i}^{n} W_{k}^{(A,B)}$ for $i = 1, \ldots, n$. Finally, for any $\sigma \in \mathfrak{S}_n$ and $k \in [n]$ we set $A_{n}^{\sigma} := \{(\sigma(1), \ldots, \sigma(k - 1))\}$ and $B_{k}^{\sigma} := \{\sigma(k + 1), \ldots, \sigma(n)\}$. Now, since $\widehat{S}_n^{(\sigma(1),B_{k}^{\sigma})} = \widehat{S}_n^{(\sigma(k),B_{k}^{\sigma})}$ for every $k \in [n]$, we obtain
\[
\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} A_{n}^{\sigma}[W] \lesssim n^{-\frac{m}{2}} \sum_{\sigma \in \mathfrak{S}_n} \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{\sigma(k) = i}^{d} \sum_{j_1, \ldots, j_m=1}^{d} E \left[ Y_{j_1, \ldots, j_m} \left( \widehat{S}_n^{(\sigma(1),B_{k}^{\sigma})} + n^{-\frac{1}{2}} \xi_{k,i} - y \right) \right] E[|W_{i,j_1}|^m] 
\]
\[
= n^{-\frac{m}{2}} \sum_{k=1}^{n} \sum_{i=1}^{n} \sum_{\sigma \in \mathfrak{S}_n} \sum_{\sigma(k) = i}^{d} \sum_{j_1, \ldots, j_m=1}^{d} E \left[ Y_{j_1, \ldots, j_m} \left( \widehat{S}_n^{(\sigma(1),B_{k}^{\sigma})} + n^{-\frac{1}{2}} \xi_{k,i} - y \right) \right] E[|W_{i,j_1}|^m] 
\]
where we use the identity $\#\{\sigma \in \mathfrak{S}_n : A_{n}^{\sigma} = A, \sigma(k) = i\} = (k - 1)!(n - k)!$ to deduce the last equality. Now, for $(A, B) \in A_{k,i}$, we have $\widehat{S}_n^{(A,B)}(i) + n^{-\frac{1}{2}} Y_i = S_n^{(A \cup \{i\},B)}$ and $\widehat{S}_n^{(A,B)}(i) + n^{-\frac{1}{2}} Y_i = S_n^{(A,B \cup \{i\})}$, so we obtain
\[
\frac{1}{n!} \sum_{i=0}^{n} \sum_{k=1}^{n} A_i^k \left[ W \right] \\
\leq n^{-\frac{m}{2}} \sum_{k=1}^{n} \frac{(k-1)(n-k)!}{n!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \\
\sum_{j_1, \ldots, j_m=1}^{d} \left\{ \frac{k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \right\} \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
= n^{-\frac{m}{2}} \sum_{k=1}^{n} \frac{k(n-k)}{(n+1)!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \sum_{j_1, \ldots, j_m=1}^{d} \\
E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
= n^{-\frac{m}{2}} \sum_{k=1}^{n} \frac{k(n-k)}{(n+1)!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \sum_{j_1, \ldots, j_m=1}^{d} \\
E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
= n^{-\frac{m}{2}} \sum_{k=0}^{n-1} \frac{k(n-k)}{(n+1)!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \sum_{j_1, \ldots, j_m=1}^{d} \\
E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
= n^{-\frac{m}{2}} \sum_{k=0}^{n-1} \frac{k(n-k)}{(n+1)!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \sum_{j_1, \ldots, j_m=1}^{d} \\
E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
= n^{-\frac{m}{2}} \sum_{k=0}^{n-1} \frac{k(n-k)}{(n+1)!} \sum_{i=1}^{d} \sum_{j(\alpha, \beta) \in \mathcal{A}_k} \sum_{j_1, \ldots, j_m=1}^{d} \\
E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
+ \frac{n+1-k}{n+1} E \left[ \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right] \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \\
\leq \sum_{i=1}^{d} \left[ E\left[ \left| W_{j_1} \right|^{m} \right] \right] \max_{1 \leq i \leq d} \left[ \left| \gamma_{j_1, \ldots, j_m} \left( \tilde{S}_{\alpha \beta j_1, \ldots, j_m}^{\mathrm{A}(1)} \right) - y_{j_1} \right| \right].
\]

Therefore, noting \(\#A_k = n! / \{k!(n-k)\}\), we conclude by Lemma 4.3 that
\[
\frac{1}{n!} \sum_{\sigma \in S_n} \sum_{k=1}^n A_\sigma^k [W] \lesssim_m n^{-\frac{m}{2}} \max_{1 \leq l \leq m} \beta^{m-l} \|h^{(l)}\|_\infty \max_{1 \leq i \leq d} \sum_{i=1}^n E[|W_{ij}|^m]. \tag{17}
\]
Next we estimate \(B_\sigma^k [W]\). Using the independence of \(X_{\sigma(k)}\) and \(Y_{\sigma(k)}\) from \(\widetilde{S}_n^\sigma(k)\) as well as Lemma 4.3, we obtain
\[
B_\sigma^k [W] \leq n^{-\frac{m}{2}} \sum_{j_1, \ldots, j_m = 1}^d E \left[ \left| Y_{j_1} \ldots Y_{j_m} \left( \widehat{S}_n^\sigma(k) - y \right) \right| \right] P \left( \mathcal{E}_{\sigma, k}^\sigma \right) \max_{1 \leq l \leq d} E[|W_{\sigma(k), l}|^m] \lesssim_m n^{-\frac{m}{2}} \max_{1 \leq l \leq m} \beta^{m-l} \|h^{(l)}\|_\infty P \left( \mathcal{E}_{\sigma, k}^\sigma \right) \max_{1 \leq l \leq d} E[|W_{\sigma(k), l}|^m].
\]
Thus, we conclude that
\[
\frac{1}{n!} \sum_{\sigma \in S_n} \sum_{k=1}^n B_\sigma^k [W] \lesssim_m n^{-\frac{m}{2}} \max_{1 \leq l \leq m} \beta^{m-l} \|h^{(l)}\|_\infty \sum_{i=1}^n P \left( \|X_i\|_\infty^\varphi \lor \|Y_i\|_\infty^\varphi > \sqrt{n/\beta} \right) \max_{1 \leq l \leq d} E[|W_{ij}|^m]. \tag{18}
\]
Combining (16) with (17, 18), we obtain
\[
\frac{1}{n!} \sum_{\sigma \in S_n} \sum_{k=1}^n L_\sigma^k [W] \lesssim_m n^{-\frac{m}{2}} \max_{1 \leq l \leq m} \beta^{m-l} \|h^{(l)}\|_\infty \left\{ \max_{1 \leq l \leq d} \sum_{i=1}^n E[|W_{ij}|^m] + \sum_{i=1}^n P \left( \|X_i\|_\infty^\varphi \lor \|Y_i\|_\infty^\varphi > \sqrt{n/\beta} \right) \max_{1 \leq l \leq d} E[|W_{ij}|^m] \right\}. \tag{19}
\]
Now, combining (12, 13) with (14, 15) and (19), we obtain the desired result. \(\square\)

4.3 Stein kernel

**Definition 4.1** (Stein kernel) Let \(F\) be a centered random vector in \(\mathbb{R}^d\). A \(d \times d\) matrix-valued measurable function \(\tau_F = (\tau_F^{ij})_{1 \leq i, j \leq d}\) on \(\mathbb{R}^d\) is called a Stein kernel for (the law of) \(F\) if \(E[|\tau_F^{ij}(F)|] < \infty\) for any \(i, j \in [d]\) and
\[
\sum_{j=1}^d E[\partial_j \varphi(F) F_j] = \sum_{i,j=1}^d E[\partial_i \varphi(F) \tau_F^{ij}(F)]
\]
for any \(\varphi \in C_b^\infty(\mathbb{R}^d)\).
When a random vector $F$ has a Stein kernel, it serves for obtaining a good upper bound of $\rho_{h,\beta}(F, Z)$ for a Gaussian vector $Z$. This is formally developed in Section 4 of Koike (2019b) with inspired by arguments in Chernozhukov et al. (2015) and Koike (2019a).

**Lemma 4.5** (Koike 2019b, Lemma 4.1) Let $F$ and $Z$ be centered random vectors in $\mathbb{R}^d$. Assume $Z$ is Gaussian. Assume also that $F$ has a Stein kernel $\tau_F = \tau_F^{ij}_{1\leq i,j \leq d}$. Then we have

$$\rho_{h,\beta}(F, Z) \leq \frac{3}{2} \max \{ \|h''\|_\infty, \beta \|h'\|_\infty \} \mathbb{E} \left[ \max_{1 \leq i,j \leq d} |\tau_F^{ij}(F) - \mathbb{E}[Z_iZ_j]| \right]$$

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

The following simple lemma plays a key role in our arguments.

**Lemma 4.6** Let $\xi = (\xi_i)_{i=1}^n$ be a sequence of independent centered random variables with unit variance and $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq d}$ be a $n \times d$ matrix. Define the $d$-dimensional random vector $F$ by

$$F_j = \sum_{i=1}^n a_{ij}\xi_i, \quad j = 1, \ldots, d.$$ 

Suppose that $\xi_i$ has a Stein kernel $\tau_i$ for every $i$ and define the $d \times d$ matrix-valued function $\tau_F = (\tau_F^{jk})_{1 \leq j,k \leq d}$ on $\mathbb{R}^d$ by

$$\tau_F^{jk}(x) = \mathbb{E} \left[ \sum_{i=1}^n a_{ij}a_{ik} \tau_i(\xi_i) \mid F = x \right], \quad x \in \mathbb{R}^d.$$ 

Then $\tau_F$ is a Stein kernel for $F$. Moreover, it holds that

$$\mathbb{E} \left[ \max_{1 \leq i,j,k \leq d} \left| \tau_F^{jk}(F) - \sum_{i=1}^n a_{ij}a_{ik} \tau_i(\xi_i) \right| \right] \leq \sqrt{2 \log(2d^2)} \max_{1 \leq i \leq d} \sqrt{\sum_{i=1}^n a_{ij}^4(\|\tau_i\|_\infty + 1)^2}. \quad (20)$$

**Proof** First, we show that $\tau_F$ is a Stein kernel for $F$. Take $\varphi \in C_b^\infty(\mathbb{R}^d)$ arbitrarily. For every $j \in [d]$, we define the function $f_j : \mathbb{R}^n \to \mathbb{R}$ by

$$f_j(x_1, \ldots, x_n) = \partial_j \varphi \left( \sum_{i=1}^n a_{ij}x_i, \ldots, \sum_{i=1}^n a_{id}x_i \right), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$ 

Also, we denote by $\mathcal{L}_i$ the law of $\xi_i$ for every $i \in [n]$. Then we have
\[ \sum_{j=1}^{d} \mathbb{E}[\partial_j \varphi(F)F_j] = \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij} \mathbb{E}[f_j(\xi_1, \ldots, \xi_n)\xi_i] \]
\[ = \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij} \int_{\mathbb{R}^n} f_j(x_1, \ldots, x_n)x_i L_1(dx_1) \cdots L_n(dx_n) (\because \text{independence of } \xi) \]
\[ = \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij} \int_{\mathbb{R}^n} \frac{\partial f_j}{\partial x_i}(x_1, \ldots, x_n)\tau_i(x_i) L_1(dx_1) \cdots L_n(dx_n) (\because \text{definition of } \tau_i) \]
\[ = \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij} E \left[ \sum_{k=1}^{d} a_{ik} \partial_{jk} \varphi(F) \tau_i(\xi_i) \right] = \sum_{j,k=1}^{d} E[\partial_{jk} \varphi(F) \tau_j^k(F)]. \]

This implies that \( \tau_F \) is a Stein kernel for \( F \).

Next we prove (20). It suffices to consider the case \( \max_{1 \leq i \leq n} \| \tau_i \|_{\infty} < \infty \). Then since \( \tau_1(\xi_1), \ldots, \tau_n(\xi_n) \) are independent, Hoeffding’s inequality (Bühlmann and van de Geer 2011, Lemma 14.14) yields

\[ E \left[ \max_{1 \leq i,j \leq d} \left| \tau_j^k(F) - \sum_{i=1}^{n} a_{ij} d_{ik} \right| \right] = E \left[ \max_{1 \leq i,j \leq d} \left| \sum_{i=1}^{n} a_{ij} d_{ik} (\tau_i(\xi_i) - 1) \right| \right] \]
\[ \leq \sqrt{2 \log(2d^2)} \max_{1 \leq i,j \leq d} \sqrt{\sum_{i=1}^{n} a_{ij}^2 d_{ik}^2 (\| \tau_i \|_{\infty} + 1)^2} \]
\[ = \sqrt{2 \log(2d^2)} \max_{1 \leq i,j \leq d} \sqrt{\sum_{i=1}^{n} a_{ij}^2 (\| \tau_i \|_{\infty} + 1)^2}. \]

This completes the proof. \( \square \)

The following estimate is a simple consequence of Nemirovski’s inequality.

**Lemma 4.7** Under the assumptions of Lemma 4.1,

\[ E \left[ \max_{1 \leq i,j,k \leq d} \frac{1}{n} \sum_{i=1}^{n} (X_{ij}X_{ik} - E[X_{ij}X_{ik}]) \right] \leq \sqrt{\frac{8 \log(2d^2)}{n}} \Delta_{n,1}. \]

**Proof** Nemirovski’s inequality (Bühlmann and van de Geer 2011, Lemma 14.24) implies that

\[ E \left[ \max_{1 \leq i,j,k \leq d} \sum_{i=1}^{n} (X_{ij}X_{ik} - E[X_{ij}X_{ik}]) \right] \leq \sqrt{8 \log(2d^2)} E \left[ \max_{1 \leq i,j \leq d} \sum_{i=1}^{n} X_{ij}^4 \right]. \]

Now the desired result follows from the Lyapunov inequality. \( \square \)
Now, we can prove Proposition 4.1 using the results established in this subsection.

**Proof of Proposition 4.1** By (Ley et al. 2017, Example 4.9(c)), \( \eta_i \) has the Stein kernel \( \tau_i^0(x) := 2^{-1} x(1-x) 1_{(0,1)}(x), \ x \in \mathbb{R} \). Then a simple computation shows that \( \xi_i \) has the Stein kernel \( \tau_i(x) := 16 \tau_i^0((x+1)/4) = 2^{-1}(x+1)(3-x) 1_{(-1,3)}(x), \ x \in \mathbb{R} \). Therefore, Lemmas 4.5 and 4.6 yield

\[
\begin{align*}
\mathbb{E} \left[ \sup_{y \in \mathbb{R}^d} \mathbb{E} \left[ h(\Phi_\rho(S_n^X - y)) \mid X \right] - \mathbb{E} \left[ h(\Phi_\rho(Z - y)) \right] \right] \\
&\leq \frac{3}{2} \left( \max_{1 \leq l \leq 2} \beta^{2-l} ||h^{(l)}||_\infty \right) \left\{ \mathbb{E} \left[ \max_{1 \leq j, k \leq d} \left| \frac{1}{n} \sum_{i=1}^n X_{ij} X_{ik} - \mathcal{C}_{jk} \right| \right] \\
&\quad + \frac{3 \sqrt{2 \log(2d^2)}}{2n} \mathbb{E} \left[ \max_{1 \leq j \leq d} \left| \sum_{i=1}^n |X_{ij}|^4 \right| \right] \right\}.
\end{align*}
\]

Hence, Lemma 4.7 and the Lyapunov inequality imply that

\[
\rho_{h,\rho}(S_n^Y, Z) \lesssim \left( \max_{1 \leq l \leq 2} \beta^{2-l} ||h^{(l)}||_\infty \right) \left\{ \Delta_{n,0}^X + \sqrt{\frac{\log d}{n}} \Delta_{n,1}^X \right\}.
\]

This completes the proof. \( \square \)

### 4.4 Proof of Lemma 4.1

**Lemma 4.8** Under the assumptions of Lemma 4.1, there is a universal constant \( C > 0 \) such that

\[
\begin{align*}
\rho_{h,\rho}(S_n^X, Z) &\leq C \left[ \left( \max_{1 \leq l \leq 2} \beta^{2-l} ||h^{(l)}||_\infty \right) \left( \Delta_{n,0}^X + \sqrt{\frac{\log d}{n}} \Delta_{n,1}^X \right) \right. \\
&\quad + \left( \max_{1 \leq l \leq 4} \beta^{4-l} ||h^{(l)}||_\infty \right) \frac{\left( \Delta_{n,1}^X \right)^2 + \Delta_{n,2}^X (\beta^{-1} \log d)}{n} \right]
\end{align*}
\]

for any \( h \in C^\infty_b(\mathbb{R}) \) and \( \beta > 0 \).
Proof Let us define $(\xi_i)_{i=1}^n$ and $Y = (Y_i)_{i=1}^n$ as in Proposition 4.1. A straightforward computation shows $E[\xi_i] = 0$ and $E[\xi_i^2] = E[\xi_i^3] = 1$, so we can apply Lemma 4.4 to $X$ and $Y$ with $m = 4$. Then noting that $|\xi_i| \leq 3$, we obtain

$$\rho_{h, \beta}(S_n^X, S_n^Y) \leq \frac{1}{n^2} \left( \max_{1 \leq l \leq 4} \beta^{4-l} \|h^{(l)}\|_\infty \right) \left\{ \max_{1 \leq j \leq d} \sum_{i=1}^n E[|X_{ij}|^4] \right. $$

$$+ \sum_{i=1}^n E[|X_i|^4 ; |X_i| > \sqrt{n}/(3\beta)] $$

$$+ \sum_{i=1}^n P\left( |X_i| > \sqrt{n}/(3\beta) \right) \max_{1 \leq j \leq d} E[|X_{ij}|^4] \right\}.$$

We evidently have

$$\frac{1}{n^2} \max_{1 \leq j \leq d} \sum_{i=1}^n E[|X_{ij}|^4] \leq \frac{\left( \Delta_{n,1}^X \right)^2}{n}.$$

Moreover, Chebyshev’s association inequality (see, e.g. Theorem 2.14 in Boucheron et al. 2013) yields

$$\sum_{i=1}^n P\left( |X_i| > \sqrt{n}/(3\beta) \right) \max_{1 \leq j \leq d} E[|X_{ij}|^4]$$

$$\leq \sum_{i=1}^n \max_{1 \leq j \leq d} E[|X_{ij}|^4 ; |X_i| > \sqrt{n}/(3\beta)].$$

So we conclude that

$$\rho_{h, \beta}(S_n^X, S_n^Y) \leq \left( \max_{1 \leq l \leq 4} \beta^{4-l} \|h^{(l)}\|_\infty \right) \left( \frac{\Delta_{n,1}^X}{n} \right)^2 + \Delta_{n,2}^X (\beta^{-1} \log d) \frac{1}{n}.$$

(21)

Since $\rho_{h, \beta}(S_n^X, Z) \leq \rho_{h, \beta}(S_n^X, S_n^Y) + \rho_{h, \beta}(S_n^Y, Z)$, we obtain the desired result from (21) and Proposition 4.1.

Proof of Lemma 4.1 Without loss of generality, we may assume

$$e^{-2} \Delta_{n,1}^X \sqrt{(\log d)^3 \frac{1}{n}} \leq 1$$

(22)

since otherwise the claim obviously holds true with $C = 1$. 
Set $\beta = \varepsilon^{-1} \log d$ (hence, $\beta^{-1} \log d = \varepsilon$). By (10), we have

$$P \left( \max_{1 \leq j \leq d} (S_{n,j}^X - y_j) \in A \right) \leq P(\Phi_{\beta}(S_n^X - y) \in A^\varepsilon) = E[1_{A^\varepsilon}(\Phi_{\beta}(S_n^X - y))].$$

Next, by Lemma 4.2 there is a $C^\infty$ function $h : \mathbb{R} \to \mathbb{R}$ and a universal constant $K > 0$ such that $\|h^{(r)}\|_\infty \leq K\varepsilon^{-r}$ for $r = 1, 2, 3, 4$ and $1_{A^\varepsilon}(x) \leq h(x) \leq 1_{A^\delta}(x)$ for all $x \in \mathbb{R}$. Then we have $E[1_{A^\varepsilon}(\Phi_{\beta}(S_n^X - y))] \leq E[h(\Phi_{\beta}(S_n^X - y))]$. Now, by Lemma 4.8 we have

$$\rho_{h,\beta}(S_n^X, Z) \lesssim \varepsilon^{-2}(\log d) \left( \Delta_{n,0}^X + \Delta_{n,1}^X \sqrt{\frac{\log d}{n}} \right) + \varepsilon^{-4}(\log d)^3 \frac{\Delta_{n,1}^X}{n} + \Delta_{n,2}^X(\varepsilon).$$

Hence, (22) yields

$$\rho_{h,\beta}(S_n^X, Z) \lesssim \varepsilon^{-2}(\log d) \left( \Delta_{n,0}^X + \Delta_{n,1}^X \sqrt{\frac{\log d}{n}} \right) + \varepsilon^{-4}(\log d)^3 \frac{\Delta_{n,1}^X}{n} + \Delta_{n,2}(\varepsilon).$$

Meanwhile, we also have

$$E[h(\Phi_{\beta}(Z-y))] \leq E[1_{A^\delta}(\Phi_{\beta}(Z-y))] \leq E \left[ 1_{A^\delta} \left( \max_{1 \leq j \leq d} (Z_j - y_j) \right) \right] = P \left( \max_{1 \leq j \leq d} (Z_j - y_j) \in A^{5\varepsilon} \right).$$

Consequently, we complete the proof. \qed

## 5 Proofs for Sect. 2

For $d$-dimensional random vector $F$, we define the $2d$-dimensional random vector $F^\circ$ by

$$F^\circ := (F_1, \ldots, F_d, -F_1, \ldots, -F_d)^\top.$$

Also, for a sequence $X = (X_i)_{i=1}^n$ of random vectors in $\mathbb{R}^d$, we set $X^\circ := (X_i^\circ)_{i=1}^n$. Note that we have $(S_n^X)^\circ = S_n^{X^\circ}$.

### 5.1 Preliminary lemmas

**Lemma 5.1** Let $Z$ be a Gaussian vector in $\mathbb{R}^d$. Then $C_{Z^\circ}(\varepsilon) \leq 2C_Z(\varepsilon)$ for any $\varepsilon > 0$.

**Proof** Take $y \in \mathbb{R}^{2d}$ arbitrarily and set $U := \max_{1 \leq j \leq d}(Z_j - y_j)$ and $V := \max_{1 \leq j \leq d}(-Z_j - y_{d+j})$. Then we have

Springer
\[ P\left(0 \leq \max_{1 \leq j \leq 2d} (Z^o_j - y_j) \leq \varepsilon\right) = P\left(\{U \lor V \geq 0\} \cap \{U \lor V \leq \varepsilon\}\right) \]
\[ \leq P(\{U \geq 0\} \cap \{U \lor V \leq \varepsilon\}) + P(\{V \geq 0\} \cap \{U \lor V \leq \varepsilon\}) \]
\[ \leq P(0 \leq U \leq \varepsilon) + P(0 \leq V \leq \varepsilon) \leq C_2(\varepsilon) + C_{-Z}(\varepsilon). \]

Now, since \(-Z\) has the same distribution as \(Z\), we have \(C_{-Z}(\varepsilon) = C_Z(\varepsilon)\). This completes the proof. \(\square\)

**Lemma 5.2**  Let \(F, Z\) be two random vectors in \(\mathbb{R}^d\) and assume \(Z\) is Gaussian. Assume also that there are constants \(\varepsilon, \eta > 0\) such that

\[ P\left(\max_{1 \leq j \leq 2d} (F^o_j - y_j) \in A\right) \leq P\left(\max_{1 \leq j \leq 2d} (Z^o_j - y_j) \in A^c\right) + \eta \tag{23} \]

for any \(y \in \mathbb{R}^d\) and Borel set \(A \subset \mathbb{R}\). Then we have

\[ \sup_{A \in \mathcal{A}^c} |P(F \in A) - P(Z \in A)| \leq 2C_Z(\varepsilon) + \eta. \]

**Proof**  Take \(y \in \mathbb{R}^d\) arbitrarily. Then we have

\[ P(F^o \leq y) = P\left(\max_j (F^o_j - y_j) \leq 0\right) \leq P\left(\max_j (Z^o_j - y_j) \leq \varepsilon\right) + \eta \quad (\because (5.1)) \]
\[ = P\left(0 < \max_j (Z^o_j - y_j) \leq 0\right) + P\left(0 < \max_j (Z^o_j - y_j) \leq \varepsilon\right) + \eta \]
\[ \leq P(Z^o \leq y) + C_Z(\varepsilon) + \eta. \]

Meanwhile, (23) yields

\[ P\left(\max_{1 \leq j \leq 2d} (F^o_j - y_j) > 0\right) \leq P\left(\max_{1 \leq j \leq 2d} (Z^o_j - y_j) > -\varepsilon\right) + \eta, \]

so we obtain

\[ P\left(\max_{1 \leq j \leq 2d} (Z^o_j - y_j) \leq -\varepsilon\right) \leq P\left(\max_{1 \leq j \leq 2d} (F^o_j - y_j) \leq 0\right) + \eta. \]

Thus, we infer that

\[ P(Z^o \leq y) = P\left(\max_j (Z^o_j - y_j) \leq -\varepsilon\right) + P\left(-\varepsilon < \max_j (Z^o_j - y_j) \leq 0\right) \]
\[ \leq P\left(\max_j (F^o_j - y_j) \leq 0\right) + \eta + C_Z(\varepsilon) = P(F^o \leq y) + \eta + C_Z(\varepsilon). \]

So we obtain \(|P(F^o \leq y) - P(Z^o \leq y)| \leq 2C_Z(\varepsilon) + \eta\) by Lemma 5.1.
\[
\sup_{A \in \mathcal{A}_n} |P(F \in A) - P(Z \in A)| = \sup_{y \in \mathbb{R}^d} |P(F^y \leq y) - P(Z^y \leq y)|,
\]

this completes the proof. \hfill \square

**Lemma 5.3** \( C_F(\varepsilon) > 0 \) for any \( d \)-dimensional random vector \( F \) and \( \varepsilon > 0 \).

**Proof** To obtain a contradiction, assume \( C_F(\varepsilon) = 0 \). Then we have \( P(x \leq F \leq x + \varepsilon) = 0 \) for all \( x \in \mathbb{R} \). Thus, \( 1 = P(F \in \mathbb{R}) = \sum_{i=-\infty}^{\infty} P(i\varepsilon \leq F \leq (i+1)\varepsilon) = 0 \), a contradiction. \hfill \square

### 5.2 Proof of Theorem 2.1(a)

The following is a generalization of Chernozhukov et al. (2017a, Lemma C.1):

**Lemma 5.4** Let \( Y \) be a non-negative random variable such that
\[
P(Y > x) \leq Ae^{-x/B} \text{ for all } y \geq 0 \text{ and for some constants } A, B > 0.
\]
Then we have
\[
E[\mathbb{1}_{\{Y > t\}y^p}] \leq p!Ae^{-t/B}(t+B)^p \text{ for every } t \geq 0 \text{ and every positive integer } p.
\]

**Proof** A simple computation yields
\[
E[\mathbb{1}_{\{Y > t\}y^p}] = p \int_0^t P(Y > t)x^{p-1}dx + p \int_t^{\infty} P(Y > x)x^{p-1}dx
\leq Ap^p e^{-t/B} + pA \int_t^{\infty} e^{-x/B}x^{p-1}dx.
\]

By Eq.(8.352.2) in Gradshteyn and Ryzhik (2007), we have
\[
pA \int_t^{\infty} x^{p-1}e^{-x/B}dx = pAB^p \int_{t/B}^{\infty} y^{p-1}e^{-y}dy = p!AB^pe^{-t/B} \sum_{q=0}^{p-1} \frac{(t/B)^q}{q!}.
\]

Consequently, we obtain
\[
E[\mathbb{1}_{\{Y > t\}y^p}] \leq p!AB^pe^{-t/B} \sum_{q=0}^{p-1} \frac{(t/B)^q}{q!} \leq p!Ae^{-t/B}(t+B)^p.
\]

This completes the proof. \hfill \square

**Lemma 5.5** If there are constants \( B_n, \kappa_n \geq 1 \) such that
\[
\max_{j} n^{-1} \sum_{i=1}^{n} E[X_{ij}^4] \leq B_n^2 \quad \text{and} \quad \max_{i,j} |X_{ij}| \leq 2\kappa_n \text{ a.s.,}
\]
there is a universal constant \( C > 0 \) such that
\[
\Delta_{n,1}^X \leq C \left( B_n + \kappa_n^2 \sqrt{\frac{\log d}{n}} \right).
\]

**Proof** By (Chernozhukov et al. 2015, Lemma 9) and assumptions, we have

\( \square \) Springer
\[
E \left[ \max_{1 \leq i \leq n} \sum_{i=1}^{n} X_{ij}^4 \right] \lesssim \max_{1 \leq i \leq d} \sum_{i=1}^{n} E \left[ X_{ij}^4 \right] + (\log d) E \left[ \max_{1 \leq i \leq n} \max_{1 \leq j \leq d} X_{ij}^4 \right] \\
\lesssim nB_n^2 + \kappa_n^4(\log d),
\]
which yields the desired result. \qed

**Lemma 5.6** Suppose that \(\max_{i,j} \|X_{ij}\|_{\Psi_1} \leq B_n\) for some \(B_n \geq 1\). Set \(\kappa_n := 2B_n \log n\). For \(i = 1, \ldots, n\) and \(j = 1, \ldots, d\), define

\[
\hat{X}_{ij} := X_{ij} 1_{\{X_{ij} > \kappa_n\}} - E \left[ X_{ij} 1_{\{X_{ij} > \kappa_n\}} \right]
\]
and \(\hat{X} := (\hat{X}_{ij})_{i=1}^n\) with \(\hat{X}_i = (\hat{X}_{i1}, \ldots, \hat{X}_{id})^\top\). Then there is a universal constant \(C > 0\) such that

\[
\|\|S_n^X\|_{\infty, \Psi_1} \| \leq C \sqrt{\delta_{n,2}}.
\]

**Proof** For every \(p = 2, 3, \ldots\), we have by Lemma 5.4

\[
E \left[ |\hat{X}_{ij}|^p \right] \leq 2^{p-1} E \left[ |X_{ij}|^p 1_{\{X_{ij} > \kappa_n\}} \right] \leq 2^p p! e^{-\kappa_n/B_n} (\kappa_n + B_n)^p
\]
\[
= \frac{p!}{2} (2\kappa_n + 2B_n)^{p-2} \cdot \frac{8(\kappa_n + B_n)^2}{n^2},
\]
so the Bernstein inequality (van der Vaart and Wellner 1996, Lemma 2.2.11) yields

\[
P\left( \|S_n^X\| > x \right) \leq 2 \exp \left( -\frac{1}{2} \frac{x^2}{8(\kappa_n + B_n)^2/n^2 + (2\kappa_n + 2B_n)x/\sqrt{n}} \right)
\]
for all \(j \in \{1, \ldots, d\}\) and \(x \geq 0\). Therefore, by Lemma 2.2.10 in van der Vaart and Wellner (1996), we obtain

\[
\|\|S_n^X\|_{\infty, \Psi_1} \| \leq \sqrt{\frac{(2\kappa_n + 2B_n)^2(\log d)^2}{n}} + \sqrt{\frac{8(\kappa_n + B_n)^2 \log d}{n^2}} \lesssim \sqrt{\delta_{n,2}}.
\]
This completes the proof. \qed

Under the present assumptions, we can reduce Lemma 4.1 to the following form:

**Lemma 5.7** Under the assumptions of Theorem 2.1(a), there is a universal constant \(C_0 > 0\) such that

\[
P\left( \max_{1 \leq i \leq d} (S_{n,j}^X - y_j) \in A \right) \leq P\left( \max_{1 \leq i \leq d} (Z_{n,j}^X - y_j) \in A^6 \right) + C_0 \epsilon^{-2} \left( \sqrt{\delta_{n,1}} + \delta_{n,2} \right),
\]
for any \(y \in \mathbb{R}^d\), \(A \in \mathcal{B}(\mathbb{R})\) and \(\epsilon \geq 12B_n(\log n)(\log d)/\sqrt{n}.
\]
\textbf{Proof} Set $\kappa_n := 2B_n \log n$. For $i = 1, \ldots, n$ and $j = 1, \ldots, d$, define
\[ \widetilde{X}_{ij} := X_{ij} 1_{|X_{ij}| \leq \kappa_n} - E \left[ X_{ij} 1_{|X_{ij}| \leq \kappa_n} \right] \]
and set $\widetilde{X} := (\widetilde{X}_{ij})_{i,j=1}^n$ with $\tilde{X}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{id})^\top$. Note that $\max_{i,j} |X_{ij}| \leq 2\kappa_n$. Also, we evidently have
\[ P \left( \max_{1 \leq i \leq d} (S^X_{n,j} - y_j) \in A \right) \leq P \left( \max_{1 \leq i \leq d} (\tilde{S}^X_{n,j} - y_j) \in A^\epsilon \right) + P \left( \|S^X_n - \tilde{X}\|_{\infty} \geq \epsilon \right). \]
(24)
Noting $E[X_{ij}] = 0$, we have $X_{ij} - \tilde{X}_{ij} = X_{ij} 1_{|X_{ij}| > \kappa_n} - E[X_{ij} 1_{|X_{ij}| > \kappa_n}]$. Hence, Lemma 5.6 and the Markov inequality yield
\[ P \left( \|S^X_n - \tilde{X}\|_{\infty} \geq \epsilon \right) \leq \epsilon^{-2} \text{E} \left[ \|S^X_n - \tilde{X}\|_{\infty}^2 \right] \leq \epsilon^{-2} \delta_{n,2}. \]
(25)
Next, applying Lemma 4.1 to $\tilde{X}$ with $\mathcal{C} = \mathcal{C}_n^X$, we obtain
\[ P \left( \max_{1 \leq i \leq d} (\tilde{S}^X_{n,j} - y_j) \in A^\epsilon \right) \leq P \left( \max_{1 \leq i \leq d} (\tilde{S}^X_{n,j} - y_j) \in A^\epsilon \right) + C \left\{ \epsilon^{-2} \Delta_{n,0} \log d + \Delta_{n,1} \sqrt{\frac{(\log d)^3}{n}} + \epsilon^{-4} \Delta_{n,2}^2 (\epsilon) \frac{\log d^3}{n} \right\}, \]
where $C > 0$ is a universal constant. The Schwarz inequality and Lemma 5.4 yield
\[ \Delta_{n,0} = \max_{1 \leq i, j \leq d} \left| \frac{1}{n} \sum_{i=1}^{n} \left( E[\tilde{X}_{ij}\tilde{X}_{ik}] - E[X_{ij}X_{ik}] \right) \right| \leq \max_{1 \leq i, j \leq d, 1 \leq k \leq n} \left( \|X_{ij} - \tilde{X}_{ij}\|_2 \|\tilde{X}_{ik}\|_2 + \|X_{ij}\|_2 \|\tilde{X}_{ik} - \tilde{X}_{jk}\|_2 \right) \leq e^{-\kappa_n/(2B_n)} B_n^2 \log n \leq B_n^2 (\log d)(\log n^2)/n. \]
(26)
Meanwhile, applying Lemma 5.5 to $\tilde{X}$ (note that $E[X_{ij}^4] \leq E[X_{ij}^4]$), we obtain
\[ \Delta_{n,1} \sqrt{\frac{(\log d)^3}{n}} \leq B_n \sqrt{\frac{(\log d)^3}{n} + \kappa_n \frac{\log d^2}{n}} \leq \sqrt{\delta_{n,1} + \delta_{n,2}}. \]
Moreover, since $\sqrt{n\epsilon}/(3 \log d) \geq 2\kappa_n$ by assumption, we have $\Delta_{n,2}(\epsilon) = 0$. Consequently, we obtain
\[ P \left( \max_{1 \leq i \leq d} (S^X_{n,j} - y_j) \in A^\epsilon \right) \leq P \left( \max_{1 \leq i \leq d} (Z^X_{n,j} - y_j) \in A^\epsilon \right) + C' \epsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}} \right), \]
where $C' > 0$ is a universal constant. Combining this with (24, 25), we complete the proof. \qed
Proof of Theorem 2.1(a) Without loss of generality, we may assume
\[ \Theta_X^{2/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}^{1/3} \right) \leq 1. \] (27)
since otherwise the claim holds true with \( C = 1 \). Noting \( \Theta_X > 0 \) by Lemma 5.3, we set
\[ \varepsilon := 12 \Theta_X^{-1/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}^{1/3} \right). \]
Then it holds that \( \varepsilon \geq 12B_n(n\log(n)d)/\sqrt{n} \). In fact,
\[ \frac{\sqrt{n}\varepsilon}{12B_n(n\log(n)d)} \geq \Theta_X^{-1/3} \frac{\sqrt{n}}{B_n(n\log(n)d)} \cdot \frac{B_n^{2/3}(\log(n)d)^2/3(\log(n))^2/3}{n^{1/3}} = \Theta_X^{-1/3} \frac{n^{1/6}}{B_n^{1/3}(\log(n))^1/3(\log(d))^{1/3}} = \left( \Theta_X^{2/3} \delta_{n,2}^{1/3} \right)^{-1/2}, \]
so (27) implies the desired inequality. Now, since the assumptions of Theorem 2.1(a) are satisfied with replacing \( X \) by \( X^* \), we can apply Lemma 5.7 to \( X^* \) instead of \( X \). Thus, there is a universal constant \( C_0 > 0 \) such that
\[ P \left( \max_{1 \leq j \leq d} (S_{n,j} - y_j) \in A \right) \leq P \left( \max_{1 \leq j \leq d} (Z_{n,j} - y_j) \in A^{6\varepsilon} \right) + C_0 \varepsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}} \right) \]
for any \( y \in \mathbb{R}^d \) and \( A \in \mathcal{B}({\mathbb{R}}) \). Noting \( S_n^{X^*} = (S_n^X)^e \), by Lemma 5.2 we obtain
\[ \rho_n(A^e) \leq 2C_n^{X^*}(6\varepsilon) + C_0 \varepsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}} \right) \leq 12\Theta_X \varepsilon + C_0 \varepsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}} \right) \leq \Theta_X^{2/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}^{1/3} \right). \]
Thus, we complete the proof. \( \square \)

5.3 Proof of Theorem 2.1(b)

In the current situation, Lemma 4.1 can be reduced to the following form:

Lemma 5.8 Under the assumptions of Theorem 2.1(b), there is a universal constant \( C_0 > 0 \) such that
\[ P \left( \max_{1 \leq j \leq d} (S_{n,j}^X - y_j) \in A \right) \leq P \left( \max_{1 \leq j \leq d} (Z_{n,j}^X - y_j) \in A^{6\varepsilon} \right) + C_0 \varepsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}(q)} \right), \]
for any \( y \in \mathbb{R}^d \), \( A \in \mathcal{B}(\mathbb{R}) \) and \( \varepsilon \geq 6D_n(\log d)^{1-1/q}/n^{1/2-1/q} \).

**Proof** The proof is parallel to that of Lemma 5.7. Set \( \kappa_n := D_n(n/ \log d)^{1/q} \) so that

\[
\kappa_n^2 \left( \frac{\log d}{n} \right)^2 = (\log d) \frac{D^q_n}{\kappa_n^{q-2}} = \delta_{n,2}(q).
\]

For \( i = 1, \ldots, n \) and \( j = 1, \ldots, d \), define

\[
\tilde{X}_{ij} := X_{ij}1_{\{\|x\|_{\infty} \leq \kappa_n\}} - E \left[ X_{ij}1_{\{\|x\|_{\infty} \leq \kappa_n\}} \right]
\]

and set \( \tilde{X} := (\tilde{X}_{ij})_{i=1}^n \) with \( \tilde{X}_i = (\tilde{X}_{i1}, \ldots, \tilde{X}_{id})^\top \). Note that \( \max_{i,j} |\tilde{X}_{ij}| \leq 2\kappa_n \). Also, we evidently have (24) with the present notation. Moreover, noting \( \text{E}[X_{ij}] = 0 \), we have \( X_{ij} - \tilde{X}_{ij} = X_{ij}1_{\{\|x\|_{\infty} > \kappa_n\}} - E[X_{ij}1\{\|x\|_{\infty} > \kappa_n\}] \). Thus, Nemirovski’s inequality and assumptions yield

\[
\text{E} \left[ \|S_n^X - \tilde{X}\|_{\infty}^2 \right] \leq \frac{\log d}{n} E \left[ \max_{1 \leq i,j \leq d} \sum_{i=1}^n X_{ij}^2 1_{\{\|x\|_{\infty} > \kappa_n\}} \right] \leq (\log d) \frac{D^q_n}{\kappa_n^{q-2}} = \delta_{n,2}(q).
\]

Hence, the Markov inequality yield

\[
P \left( \|S_n^X - \tilde{X}\|_{\infty} \geq \varepsilon \right) \leq \varepsilon^{-2} \delta_{n,2}(q).
\]  

(28) Next, applying Lemma 4.1 to \( \tilde{X} \) with \( \mathcal{C} = \mathcal{C}_n^X \), we obtain

\[
P \left( \max_{1\leq i \leq d} (\tilde{X}_{n,j} - y_j) \in A' \right) \leq P \left( \max_{1\leq i \leq d} (Z_{n,j}^X - y_j) \in A'_\varepsilon \right)
\]

\[
+ C \left\{ \varepsilon^{-2} \left( \Delta_{n,0} \log d + \Delta_{n,1}^X \sqrt{\frac{(\log d)^3}{n}} \right) + \varepsilon^{-4} \Delta_{n,2}^X (\log d)^3 \right\},
\]

where \( C > 0 \) is a universal constant. Noting \( \text{E}[X_{ij}] = 0 \), we have

\[
\Delta_{n,0}^X := \max_{1 \leq i,j \leq d} \left| \frac{1}{n} \sum_{i=1}^n \left( \text{E}[\tilde{X}_{ij}\tilde{X}_{ik}] - \text{E}[X_{ij}X_{ik}] \right) \right|
\]

\[
= \max_{1 \leq i,j,k \leq d} \left| \frac{1}{n} \sum_{i=1}^n \text{Cov} \left[ X_{ij}1_{\{\|x\|_{\infty} > \kappa_n\}}, X_{ik}1_{\{\|x\|_{\infty} > \kappa_n\}} \right] \right|
\]

\[
\leq \max_{1 \leq i,j \leq d} \max_{1 \leq i \leq n} \text{E} \left[ X_{ij}^2 1_{\{\|x\|_{\infty} > \kappa_n\}} \right] \leq \frac{D^q_n}{\kappa_n^{q-2}}.
\]  

(29)
Meanwhile, applying Lemma 5.5 to $\widetilde{X}$ (note that $E[\widetilde{X}^2] \leq E[X^2] \mid y_j$), we obtain

$$\Delta_{n,2}^{\tilde{X}} \sqrt{\frac{(\log d)^3}{n}} \leq B_n \sqrt{\frac{(\log d)^3}{n}} + \kappa^2_n \frac{(\log d)^2}{n} = \sqrt{\delta_{n,1} + \delta_{n,2}(q)}.$$  

Moreover, since $\sqrt{n\epsilon/(3 \log d)} \geq 2\kappa_n$ by assumption, we have $\Delta_{n,2}^{\tilde{X}}(\epsilon) = 0$. Consequently, we obtain

$$P\left( \max_{1 \leq j \leq d} (\delta_{n,j}^{\tilde{X}} - y_j) \in A^\epsilon \right) \leq P\left( \max_{1 \leq j \leq d} (Z_j^X - y_j) \in A^{6\epsilon} \right) + C' \epsilon^{-2} \left( \sqrt{\delta_{n,1} + \delta_{n,2}(q)} \right),$$

where $C' > 0$ is a universal constant. Combining this with (24) and (28), we complete the proof. \hfill $\square$

**Proof of Theorem 2.1(b)** Without loss of generality, we may assume

$$\Theta_X^{2/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}(q)^{1/3} \right) \leq 1.$$  

(30)

Since otherwise the claim holds true with $K_q = 1$. Noting $\Theta_X > 0$ by Lemma 5.3, we set

$$\epsilon := 6\Theta_X^{-1/3} \left( \delta_{n,1}^{1/6} + \delta_{n,2}(q)^{1/3} \right).$$

Then it holds that $\epsilon \geq 6D_n(\log d)^{1-1/q}/n^{1/2-1/q}$. In fact,

$$\frac{n^{1/2-1/q} \epsilon}{6D_n(\log d)^{1-1/q}} \geq \Theta_X^{-1/3} \frac{n^{1/2-1/q}}{D_n(\log d)^{1-1/q}} \frac{D_n^{2/3}(\log d)^{2/3-2/(3q)}}{n^{1/3-2/(3q)}} = \Theta_X^{-1/3} \frac{n^{1/6-1/(3q)}}{D_n^{1/3}(\log d)^{1/3-1/(3q)}} = \left( \Theta_X^{2/3} \delta_{n,2}(q)^{1/3} \right)^{-1/2},$$

so (30) implies the desired inequality. Now, the remaining proof is almost the same as that of Theorem 2.1(a), where we use Lemma 5.8 instead of Lemma 5.7. \hfill $\square$

### 5.4 Proof of Lemma 2.2

Take $y \in \mathbb{R}^d$ arbitrarily. Define $A_0 := \emptyset$ and $A_j := \{Z_j + a_j\zeta - y_j \geq 0\}$ for $j = 1, \ldots, d$. We also define $B_j := A_j \setminus (A_0 \cup A_1 \cup \cdots \cup A_{j-1})$. By construction $B_1, \ldots, B_d$ are mutually exclusive and $\{\max_{1 \leq j \leq d} (Z_j + a_j\zeta - y_j) \geq 0\} = \bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} B_j$. Therefore, we have
Now, using the independence of $\zeta$ from $Z$, we deduce that

\[
P\left(0 \leq \max_{1 \leq j \leq d} (Z_j + a_j \zeta - y_j) \leq \varepsilon \right)
= \sum_{j=1}^{d} P\left(B_j \cap \left\{ \max_{1 \leq k \leq d} (Z_k + a_k \zeta - y_k) \leq \varepsilon \right\}\right)
\leq \sum_{j=1}^{d} P\left(B_j \cap \{Z_j + a_j \zeta - y_j \leq \varepsilon \}\right)
= \sum_{j=1}^{d} P\left(B_j \cap \{-(Z_j - y_j)/|a_j| \leq (a_j/|a_j|)\zeta \leq \varepsilon /|a_j| - (Z_j - y_j)/|a_j| \}\right).
\]

This yields the desired result. \(\square\)

### 5.5 Proof of Proposition 2.1

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

\[
\rho := \limsup_{n \to \infty} \left| P\left(\max_{1 \leq j \leq d} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{ij} \leq x_n\right) - P\left(\max_{1 \leq j \leq d} \zeta_j \leq x_n\right) \right| > 0.
\]

The proof of this result is a slight refinement of the arguments in Chen (2018, Remark 1). First, by Theorem 1 in Petrov (1975), Chapter VIII) (see also Eq.(2.41) in Petrov (1975, Chapter VIII)), if a sequence $x_n \geq 0$ satisfies $x_n = O(n^{1/6})$ as $n \to \infty$, we have

\[
P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i1} > x_n\right) = \exp\left(\frac{\gamma}{6\sqrt{n}} x_n^3\right) + O\left(\frac{x_n + 1}{\sqrt{n}}\right)
\] (31)

as $n \to \infty$, where $\Phi$ denotes the cumulative distribution function of the standard normal distribution.

Next, for every $n$, we define $x_n \in \mathbb{R}$ as the solution of the equation $\Phi(x)^{d_n} = e^{-1}$, i.e. $x_n := \Phi^{-1}(e^{-1/d_n})$. Then we have $x_n - \sqrt{2 \log d_n} = o(1/\sqrt{2 \log d_n})$ as $n \to \infty$. To see this, we set
\[ b_n := \sqrt{2 \log d_n} - \frac{\log \log d_n + \log 4\pi}{2 \sqrt{2 \log d_n}}. \]

Then it is well known (e.g. Embrechts et al. 1997, Eq.(3.40)) that
\[ P(\sqrt{2 \log d_n}(\max_{1 \leq j \leq d_n} \xi_j - x_n) \leq t) \to \Lambda(t) \text{ as } n \to \infty \text{ for every } t \in \mathbb{R}, \]
where \( \Lambda(t) := \exp(-e^{-t}) \). Moreover, since \( \Lambda \) is continuous, by van der Vaart (1998, Lemma 2.11) we indeed have
\[ \lim \sup_{n \to \infty} \left| P\left( \sqrt{2 \log d_n}(\max_{1 \leq j \leq d_n} \xi_j - b_n) \leq t \right) - \Lambda(t) \right| = 0. \]

Since \( \Phi(x_n)^{d_n} = P(\sqrt{2 \log d_n}(\max_{1 \leq j \leq d_n} \xi_j - b_n) \leq \sqrt{2 \log d_n}(x_n - b_n)) \), we obtain
\[ \Lambda(\sqrt{2 \log d_n}(x_n - b_n)) \to e^{-1} \text{ as } n \to \infty. \]
Since \( \Lambda^{-1}(e^{-1}) = 0 \), this implies the desired result.

Now, since \( \lim_{n \to \infty} x_n / \sqrt{2 \log d_n} = 1 \), we have \( x_n/n^{1/6} \to \sqrt{2}c^{1/6} \) as \( n \to \infty \) by assumption. Hence, (31) yields
\[ P\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i1} > x_n \right) \frac{1}{1 - \Phi(x_n)} \to \exp \left( \frac{\gamma \sqrt{2}c}{3} \right) \]

as \( n \to \infty \). Since \( \gamma < 0 \) by assumption, there is a constant \( a \in (0, 1) \) such that
\[ P\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i1} > x_n \right) \frac{1}{1 - \Phi(x_n)} \leq a \]
for sufficiently large \( n \). For such an \( n \), we obtain
\[ P\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i1} \leq x_n \right) \geq 1 - a(1 - \Phi(x_n)) = \Phi(x_n) + (1 - a)(1 - \Phi(x_n)). \]

Now we infer that
\[ \rho = \lim \sup_{n \to \infty} \left| P\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i1} \leq x_n \right)^{d_n} \Phi(x_n)^{d_n} - 1 \right| \]
\[ \geq \lim \sup_{n \to \infty} \Phi(x_n)^{d_n} \left( 1 + (1 - a)(1 - \Phi(x_n))/\Phi(x_n) \right)^{d_n} - 1 \]
\[ \geq \lim \sup_{n \to \infty} \Phi(x_n)^{d_n} \cdot d_n(1 - a)(1 - \Phi(x_n))/\Phi(x_n), \]
where the last inequality follows from the inequality \( (1 + t)^{d_n} \geq 1 + d_n t \) holding for all \( t \geq 0 \). Since \( \Phi \) is bounded by 1, we obtain
\[ \rho \geq \lim \sup_{n \to \infty} \Phi(x_n)^{d_n} \cdot d_n(1 - a)(1 - \Phi(x_n)) = \frac{1 - a}{e} \lim \sup_{n \to \infty} d_n(1 - \Phi(x_n)). \]
Since \( d_n(1 - \Phi(x_n)) = -(e^{-1/d_n} - 1)/(1/d_n) \to 1 \) as \( n \to \infty \), we conclude \( \rho \geq (1 - a)/e > 0 \).

\[ \square \]

\section{6 Proofs for Sect. 3}

Throughout this section, we use the following notation: we set \( Y = (Y_i)_{i=1}^n := (X_i - \bar{X})_{i=1}^n \). Given a sequence \( \xi = (\xi_i)_{i=1}^n \) of random vectors, we set \( w^\xi := (w_i^\xi)_{i=1}^n \). Note that we have \( S_{n}^{WB} = S_{n}^{nY} \).

\subsection{6.1 Proof of Theorem 3.1(a)}

We may assume \( \Theta_X^{2/3} \left( (b^2 \delta_{n,1})^{1/6} + (b^2 \delta_{n,2})^{1/3} \right) \leq 1 \) without loss of generality. Set

\[ \varepsilon := 24 \Theta_X^{-1/3} \left( (b^2 \delta_{n,1})^{1/6} + (b^2 \delta_{n,2})^{1/3} \right) \]

and \( \kappa_n := 2B_n \log n \). As in the proof of Theorem 2.1(a), we can prove \( \varepsilon \geq 3b \cdot 4\kappa_n (\log d)/\sqrt{n} \). Now, we define \( \bar{X} = (\bar{X}_i)_{i=1}^n \) as in the proof of Lemma 5.7. Then we set \( Y = (Y_i)_{i=1}^n := (\bar{X}_i - \bar{X})_{i=1}^n \). Note that \( \max_{i,j} \|\bar{Y}_{ij}\| \leq 4\kappa_n \), so we have \( \max_{i,j} \|w_i^\xi \bar{Y}_{ij}\| \leq b \cdot 4\kappa_n \). We apply Lemma 4.1 to \( w^\xi \bar{Y} \) with \( C = E[S_n^X (S_n^X)^\top] \), conditionally on \( X \). Then we conclude that there is an event \( \Omega_0 \in \mathcal{F} \) such that \( P(\Omega_0) = 1 \) and

\[ P \left( \max_{1 \leq i \leq 2d} \left( S_{n,i}^{\bar{Y}} - y_j \right) \in A \mid X \right) \leq P \left( \max_{1 \leq i \leq 2d} \left( Z_{n,i}^{X} - y_j \right) \in A^{6\varepsilon} \right) \]

\[ + C\varepsilon^{-2} \left( \Delta_{n,0}^* \log d + \Delta_{n,1}^* \sqrt{\frac{(\log d)^3}{n}} \right) \]

on \( \Omega_0 \)

for any \( y \in \mathbb{R}^{2d} \) and \( A \in \mathcal{B}(\mathbb{R}) \), where \( C > 0 \) is a universal constant and

\[ \Delta_{n,0}^* := E \left[ \max_{1 \leq i \leq d} \left( \frac{1}{n} \sum_{i=1}^n \left( w_i^2 \bar{Y}_{ij} \bar{Y}_{ik} - E[X_{ij} X_{ij}] \right) \mid X \right) \right], \]

\[ \Delta_{n,1}^* := \sqrt{\frac{1}{n} E \left[ \max_{1 \leq i \leq d} \sum_{i=1}^n w_i^4 \bar{Y}_{ij}^4 \mid X \right]}. \]

Thus, we obtain

\[ P \left( \max_{1 \leq i \leq 2d} \left( \left( S_{n,i}^{WB} \right)^{\circ} - y_j \right) \in A \mid X \right) \leq P \left( \max_{1 \leq i \leq 2d} \left( Z_{n,i}^{X} - y_j \right) \in A^{6\varepsilon} \right) \]

\[ + P \left( \| S_{n}^{\circ (1 - \bar{Y}_{ij}^\circ)} \|_{\infty} > \varepsilon \mid X \right) + C\varepsilon^{-2} \left( \Delta_{n,0}^* \log d + \Delta_{n,1}^* \sqrt{\frac{(\log d)^3}{n}} \right) \]

on \( \Omega_0 \)

(32)
for any \( y \in \mathbb{R}^d \) and \( A \in B(\mathbb{R}) \). Now, noting that \( w_i \)'s are bounded by \( b \) and \( \sqrt{n}(\tilde{X} - \bar{X}) = S_n^{\tilde{X} - \bar{X}} \), the same argument as in the proof of (25) yields

\[
P\left( \| S_n^{wY - \bar{Y}} \|_{\ell_\infty} > \epsilon \right) \lesssim \epsilon^{-2} \frac{b^2 B^2_n (\log d)^2 (\log n)^2}{n}.
\]

Meanwhile, Lemmas 4.7 and 5.5 and the inequality \( E[w_i^4] \leq b^2 E[w_i^2] = b^2 \) imply that

\[
E \left[ \max_{1 \leq j,k \leq d} \left| \frac{1}{n} \sum_{i=1}^{n} \left( w_i^2 \tilde{Y}_{ij} \tilde{Y}_{ik} - E \left[ w_i^2 \tilde{X}_{ij} \tilde{X}_{ik} \right] \right) \right| \right] 
\lesssim \sqrt{\frac{\log d}{n}} \Delta_{n,1}^{w\tilde{X}} + E \left[ \| \tilde{X} \|_{\ell_\infty}^2 \right] E \left[ \frac{1}{n} \sum_{i=1}^{n} w_i^2 \right] 
\lesssim \sqrt{\frac{b^2 B^2_n \log d}{n} + \frac{b^2 \kappa_n^2 \log d}{n}} + E \left[ \| \tilde{X} \|_{\ell_\infty}^2 \right].
\]

Also, \( E \left[ \| \tilde{X} \|_{\ell_\infty}^2 \right] \lesssim \kappa_n^2 (\log d) / n \) by Lemma 14.14 in Bühlmann and van de Geer (2011). Then since \( E[w_i^4 \tilde{X}_{ij} \tilde{X}_{ij}] = E[\tilde{X}_{ij} \tilde{X}_{ij}] \) and \( b \geq 1 \), the above inequalities and (26) yield

\[
E[\Delta_{n,0}^*] \lesssim \sqrt{\frac{b^2 B^2_n \log d}{n} + \frac{b^2 B^2_n (\log d) (\log n)^2}{n}}.
\]

Moreover, since the Jensen inequality yields \( \tilde{X}_j \leq n^{-1} \sum_{i=1}^{n} \tilde{X}_{ij}^4 \), we have \( E[\Delta_{n,1}^*] \lesssim \Delta_{n,1}^{w\tilde{X}} + b \Delta_{n,1}^{\tilde{X}} \) by the Lyapunov inequality. Thus, Lemma 5.5 implies that

\[
E[\Delta_{n,1}^*] \sqrt{\frac{(\log d)^3}{n}} \lesssim \sqrt{\frac{b^2 B^2_n \log d}{n} + \frac{b^2 B^2_n (\log d) (\log n)^2}{n}}.
\]

Combining these estimates with Lemma 5.2, we obtain

\[
E [\rho_n^{WB}(A^c_e(d))] \leq 2C_{Z_n^e}^{\delta}(6\epsilon) + C_1 \epsilon^{-2} \left( \sqrt{b^2 \delta_{n,1} + b^2 \delta_{n,2}} \right),
\]

where \( C_1 > 0 \) is a universal constant. Since \( C_{Z_n^e}^{\delta}(6\epsilon) \leq 6\Theta X \epsilon \) by definition, we obtain the desired result by the definition of \( \epsilon \).

\[
6.2 \ \text{Proof of Theorem 3.1(b)}
\]

The proof is completely parallel to that of Theorem 3.1(a), where we suitably modify the definitions of \( \epsilon, \kappa_n, \tilde{X} \) and consider (28, 29) instead of (25, 26), respectively. The detail is omitted.
Acknowledgements The author thanks Denis Chetverikov and Arun Kumar Kuchibhotla for their valuable comments. The author is also grateful to two anonymous referees for their constructive comments which have significantly improved a former version of this paper. This work was supported by JST CREST Grant number JPMJCR14D7 and JSPS KAKENHI Grant numbers JP17H01100, JP18H00836, JP19K13668.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest.

Appendix: Proofs for Proposition 4.2 and Corollary 4.1 via the standard Lindeberg method

The following lemma is a counterpart of Lemma 4.4 but based on the non-randomized Lindeberg method.

Lemma A.1 Under the assumptions of Lemma 4.4, we have

\[
\rho_{h,p}(S_n^X, S_n^Y) \leq C'_m n^{-\frac{m}{2}} \left( \max_{1 \leq i \leq m} \beta^{m-1} \|h^{(i)}\|_{\infty} \right) \left\{ \sum_{i=1}^{n} E[\|X_i\|_m^m] + \sum_{i=1}^{n} E[\|Y_i\|_m^m] \right\},
\]

where \( C'_m > 0 \) depends only on \( m \).

Proof We may assume that \( X \) and \( Y \) are independent without loss of generality. Take a vector \( y \in \mathbb{R}^d \) and define the function \( \Psi : \mathbb{R}^d \to \mathbb{R} \) by \( \Psi(x) = h(\Phi_p(x - y)) \) for \( x \in \mathbb{R}^d \). For every \( k \in \{n\} \), we set

\[
S_n(k) := \frac{1}{\sqrt{n}} \sum_{i=1}^{k} X_i + \frac{1}{\sqrt{n}} \sum_{i=k+1}^{n} Y_i \quad \text{and} \quad \hat{S}_n(k) := S_n(k) - \frac{1}{\sqrt{n}} X_k.
\]

By construction \( \hat{S}_n(k) \) is independent of \( X_k \) and \( Y_k \). We also have \( S_n(k) = \hat{S}_n(k) + n^{-1/2} X_k \) and \( S_n(k-1) = \hat{S}_n(k) + n^{-1/2} Y_k \) (with \( S_n(0) := n^{-1/2} \sum_{i=1}^{n} Y_i \)). Moreover, it holds that \( S_n(n) = S_n^X \) and \( S_n(0) = S_n^Y \). Therefore, we have

\[
\left| E \left[ \Psi(S_n^X) \right] - E \left[ \Psi(S_n^Y) \right] \right| \leq \sum_{k=1}^{n} \left| E \left[ \Psi(S_n(k)) \right] - E \left[ \Psi(S_n(k-1)) \right] \right|. \tag{33}
\]

Meanwhile, when \( W = X \) or \( W = Y \), Taylor’s theorem and the independence of \( W_k \) from \( \hat{S}_n(k) \) yield

\[
E \left[ \Psi \left( \hat{S}_n(k) + n^{-1/2} W_k \right) \right] = \sum_{\lambda \in \mathbb{Z}^d : |\lambda| \leq m-1} \frac{n^{-|\lambda|/2}}{\lambda!} E \left[ \partial^\lambda \Psi \left( \hat{S}_n(k) \right) \right] E \left[ W_k^\lambda \right] + R_k(W),
\]

where
\[ R_k[W] := n^{-m/2} \sum_{\lambda \in \mathbb{Z}_+^d : |\lambda| = m} \frac{m}{\lambda!} \int_0^1 (1 - t)^{m-1} E \left[ \partial^\lambda \Psi \left( \frac{S_n(k)}{m^{1/2} W_k} \right) W_k^2 \right] dt. \]

Since we have \( E[X_i] = E[Y_i] \) for all \( i \in [N] \) and \( \lambda \in \mathbb{Z}_+^d \) with \( |\lambda| \leq m - 1 \) by assumption, we obtain

\[ \left| E \left[ \Psi \left( S_n(k) \right) \right] - E \left[ \Psi \left( S(k-1) \right) \right] \right| \leq |R_k[X]| + |R_k[Y]|. \quad (34) \]

Now, for any random vector \( W \) in \( \mathbb{R}^d \) we have by Lemma 4.3

\[ |R_k[W]| \leq c'_m n^{-\frac{m}{2}} \max_{1 \leq i \leq m} \beta^{m-i} \|h^{(i)}\|_\infty \sum_{i=1}^n E[\|W_i\|_\infty^m], \]

where \( c'_m \) depends only on \( m \). Combining this estimate with (33) and (34), we obtain the desired result. \( \square \)

**Proof of Proposition 4.2** Without loss of generality, we may assume

\[ \epsilon^{-2} \sqrt{\frac{B_0^4 (\log d)^3}{n}} \leq 1 \quad (35) \]

since otherwise the claim obviously holds true with \( C = 1 \).

Set \( \beta = \epsilon^{-1} \log d \) (hence, \( \beta^{-1} \log d = \epsilon \)). By (10), we have

\[ P \left( \max_{1 \leq j \leq d} (S_{n,j}^X - y_j) \in A \right) \leq P(\Phi_{\beta} (S_n^X - y) \in A^d) = E[1_{A^d} (\Phi_{\beta} (S_n^X - y))] \]

Next, by Lemma 4.2 there is a \( C^\infty \) function \( h : \mathbb{R} \to \mathbb{R} \) and a universal constant \( K > 0 \) such that \( \|h^{(r)}\|_\infty \leq K \epsilon^{-r} \) for \( r = 1, 2, 3, 4 \) and \( 1_{A^d}(x) \leq h(x) \leq 1_{A^d}(x) \) for all \( x \in \mathbb{R} \). Then we have \( E[1_{A^d} (\Phi_{\beta} (S_n^X - y))] \leq E[\Phi_{\beta} (S_n^X - y))]. \) Now, let us define \( Y = (Y_i^n)_{i=1} \) as in Proposition 4.1. Then we have

\[ \rho_{h,\beta} (S_n^X, S_n^Y) \lesssim \epsilon^{-4} \frac{B_0^4 (\log d)^3}{n} \]

by Lemma A.1. Combining this with Proposition 4.1, we obtain
\[ \rho_{h,\beta}(S^X_n, Z) \leq \rho_{h,\beta}(S^X_n, S^Y_n) + \rho_{h,\beta}(S^Y_n, Z) \]
\[ \leq \varepsilon^{-2}(\log d) \left( \Delta_{n,0}^X + \Delta_{n,1}^X \sqrt{\frac{\log d}{n}} \right) + \varepsilon^{-4} B_4^2 \left( \frac{\log d}{n} \right)^3 \]
\[ \leq \varepsilon^{-2}(\log d) \Delta_{n,0}^X + \varepsilon^{-2} B_2^2 \sqrt{\frac{(\log d)^3}{n}} + \varepsilon^{-4} B_4^2 \left( \frac{\log d}{n} \right)^3 \]
\[ \leq \varepsilon^{-2}(\log d) \Delta_{n,0}^X + 2\varepsilon^{-2} \sqrt{\frac{B_4^2 (\log d)^3}{n}}, \]

where the last inequality follows from (35). Meanwhile, we also have
\[ \mathbb{E}[h(\Phi_p(Z - y))] \leq \mathbb{E}[1_{A_{\text{ex}}}(\Phi_p(Z - y))] \leq \mathbb{E} \left[ 1_{A_{\text{ex}}} \left( \max_{1 \leq j \leq d} (Z_j - y_j) \right) \right] \]
\[ = P \left( \max_{1 \leq j \leq d} (Z_j - y_j) \in A_{\text{ex}} \right). \]

Consequently, we complete the proof. \(\square\)

**Proof of Corollary 4.1** Set \( \varepsilon := \Theta_X^{-1/3} (n^{-1} \log^3 d)^{1/6} \). Applying Proposition 4.2 to \( X^\circ \), we obtain
\[ P \left( \max_{1 \leq j \leq 2d} (S^X_n - y_j) \in A \right) \leq P \left( \max_{1 \leq j \leq 2d} (S^X_n - y_j) \in A_{\text{ex}} \right) + C_0 \varepsilon^{-2} \sqrt{\frac{B_4^4 \log^3 d}{n}}, \]
where \( C_0 > 0 \) is a universal constant. Since \( S^X_n = (S^X_n)^\circ \), we obtain by Lemma 5.2
\[ \rho_n(A_{\text{ex}}) \leq 2C_{Z_n}(6\varepsilon) + C_0 \varepsilon^{-2} \sqrt{\frac{\log^3 d}{n}} \lesssim \Theta_X^{2/3} \left( \frac{B_4^4 \log^3 d}{n} \right)^{1/6}. \]

This proves the first inequality in Corollary 4.1. The second one follows from Lemma 2.1. \(\square\)

**References**

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., & Kato, K. (2018). High-dimensional econometrics and regularized GMM, working paper. Retrieved from arXiv:1806.01888.
Belloni, A., Chernozhukov, V., & Kato, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika*, 102(1), 77–94.
Belloni, A., & Oliveira, R.I. (2018). A high dimensional central limit theorem for martingales, with applications to context tree models, working paper. arXiv:1809.02741.
Bentkus, V. (2005). A Lyapunov-type bound in \( R^d \). *Theory of Probability & Its Applications*, 49(2), 311–323.
Bonis, T. (2020). Stein’s method for normal approximation in Wasserstein distances with application to the multivariate Central Limit Theorem. Probability Theory and Related Fields (forthcoming). Retrieved from arXiv:1905.13615.
Boucheron, S., Lugosi, G., & Massart, P. (2013). Concentration inequalities: A nonasymptotic theory of independence. Oxford: Oxford University Press.

Bühlmann, P., & van de Geer, S. (2011). Statistics for high-dimensional data. New York: Springer.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. The Annals of Statistics, 41(6), 2786–2819.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2014). Gaussian approximation of suprema of empirical processes. The Annals of Statistics, 42(4), 1564–1597.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Probability Theory and Related Fields, 162, 47–70.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. Stochastic Processes and their Applications, 126, 3632–3651.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2017a). Central limit theorems and bootstrap in high dimensions. Annals of Probability, 45(4), 2309–2353.

Chernozhukov, V., Chetverikov, D., & Kato, K. (2017b). Detailed proof of Nazarov’s inequality, unpublished paper. Retrieved from arXiv:1711.10696.

Courtade, T. A., Fathi, M., & Pananjady, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. Annale Institut Henri Poincaré, Probabilités et Statistiques, 55(2), 777–790.

Deng, H., & Zhang, C.H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. Annals of Statistics (forthcoming). Retrieved from arXiv:1705.09528v2.

Eldan, R., Mikulincer, D., & Zhai, A. (2020). The CLT in high dimensions: Quantitative bounds via martingale embedding. Annals of Probability, 48(5), 2494–2524.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). Modelling extremal events. New York: Springer.

Fathi, M. (2019). Stein kernels and moment maps. Annals of Probability, 47(4), 2172–2185.

Gradshteyn, I., & Ryzhik, I. (2007). Table of integrals, series, and products (7th ed.). Amsterdam: Elsevier.

Hall, P. (2006). Some contemporary problems in statistical science. In A. Quirós & F. Chamizo (Ed.), Madrid inteligencer (pp. 38–41). New York: Springer.

Koike, Y. (2019a). Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. Annals of Statistics, 47(3), 1663–1687.

Koike, Y. (2019b). High-dimensional central limit theorems for homogeneous sums, working paper. Retrieved from arXiv:1902.03809.

Kuchibhotla, A.K., Mukherjee, S., & Banerjee, D. (2019). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. Bernoulli (forthcoming). Retrieved from arXiv:1806.06133v3.

Le Cam, L. (1986). Asymptotic methods in statistical decision theory. New York: Springer.

Ley, C., Reinert, G., & Swan, Y. (2017). Stein’s method for comparison of univariate distributions. Probability Surveys, 14, 1–52.

Lopes, M. E., Lin, Z., & Müller, H. G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. Annals of Probability, 48(2), 1214–1229.

Petrov, V. V. (1975). Sums of independent random variables. New York: Springer.

van der Vaart, A. W. (1998). Asymptotic statistics. Cambridge: Cambridge University Press.

van der Vaart, A. W., & Wellner, J. A. (1996). Weak convergence and empirical processes. New York: Springer.

Zhai, A. (2018). A high-dimensional CLT in $W^2_2$ distance with near optimal convergence rate. Probability Theory and Related Fields, 170(3–4), 821–845.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.