cemmap
centre for microdata methods and practice

Central Limit Theorems and Bootstrap in High Dimensions

Victor Chernozhukov
Denis Chetverikov
Kengo Kato

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP49/14

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

CENTRAL LIMIT THEOREMS AND BOOTSTRAP IN HIGH DIMENSIONS

VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND KENGO KATO

Abstract. In this paper, we derive central limit and bootstrap theorems for probabilities that centered high-dimensional vector sums hit rectangles and sparsely convex sets. Specifically, we derive Gaussian and bootstrap approximations for the probabilities $\mathrm{P}(n^{-1/2}\sum_{i=1}^n X_i \in A)$ where $X_1,\dots,X_n$ are independent random vectors in $\mathbb{R}^p$ and A is a rectangle, or, more generally, a sparsely convex set, and show that the approximation error converges to zero even if $p = p_n \to \infty$ and $p \gg n$; in particular, p can be as large as $O(e^{Cn^c})$ for some constants $c, C > 0$. The result holds uniformly over all rectangles, or more generally, sparsely convex sets, and does not require any restrictions on the correlation among components of X_i. Sparsely convex sets are sets that can be represented as intersections of many convex sets whose indicator functions depend nontrivially only on a small subset of their arguments, with rectangles being a special case.

1. Introduction

Let $X_1,\dots,X_n$ be independent random vectors in $\mathbb{R}^p$ where $p \geq 2$ may be large or even much larger than n. Denote $X_i = (X_{i1},\dots,X_{ip})'$ for $i = 1,\dots,n$. Assume that each X_i is centered, namely $\mathrm{E}[X_{ij}] = 0$, and $\mathrm{E}[X_{ij}^2] < \infty$ for all $i = 1,\dots,n$ and $j = 1,\dots,p$. Define the normalized sum

$$S_n^X := (S_{n1}^X,\dots,S_{np}^X)' := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i. \tag{1}$$

We consider Gaussian approximation to S_n^X, and to this end, let $Y_1,\dots,Y_n$ be independent centered Gaussian random vectors in $\mathbb{R}^p$ such that each Y_i has the same covariance matrix as X_i, that is, $Y_i \sim N(0, \mathrm{E}[X_iX_i'])$. Define the normalized sum for the Gaussian random vectors:

$$S_n^Y := (S_{n1}^Y,\dots,S_{np}^Y)' := \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i. \tag{2}$$

Date: First version: February 07, 2014. This version: December 11, 2014.

Key words and phrases. Central limit theorem, bootstrap limit theorems, high dimensions, rectangles, sparsely convex sets.

We are grateful to Evarist Gine, Ramon Van Handel, Vladimir Koltchinskii, Richard Nickl, and Larry Wasserman for useful discussions.

We are interested in bounding the quantity

$$\rho_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} \left| \mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A) \right|, \tag{3}$$

where $\mathcal{A}$ is a class of (Borel) sets in $\mathbb{R}^p$.

Bounding $\rho_n(\mathcal{A})$ for various classes $\mathcal{A}$ of sets in $\mathbb{R}^p$, with a special emphasis on explicit dependence on the dimension p in bounds, has been studied by a number of authors; see, for example, [5, 6, 7, 20, 26, 31, 32, 33, 34] (see [15] for an exhaustive literature review). Typically, we are interested in how fast $p = p_n \to \infty$ is allowed to grow while guaranteeing $\rho_n(\mathcal{A}) \to 0$. In particular, Bentkus [6] established one of the sharpest results in this direction and proved that, when $X_1, \dots, X_n$ are i.i.d. with $\mathrm{E}[X_i X_i'] = I$,

$$\rho_n(\mathcal{A}) \le c_p(\mathcal{A}) \mathrm{E}[\|X_i\|^3]/\sqrt{n}, \tag{4}$$

where $c_p(\mathcal{A})$ is a constant that depends only on p and $\mathcal{A}$; for example, $c_p(\mathcal{A})$ is bounded by a universal constant when $\mathcal{A}$ is the class of all Euclidean balls in $\mathbb{R}^p$, and $c_p(\mathcal{A}) \le 400 p^{1/4}$ when $\mathcal{A}$ is the class of all convex sets in $\mathbb{R}^p$. Note, however, that this bound does not allow p to be larger than n once we require $\rho_n(\mathcal{A}) \to 0$. Indeed by Hölder's inequality, when $\mathrm{E}[X_i X_i'] = I$, $\mathrm{E}[\|X_i\|^3] \ge (\mathrm{E}[\|X_i\|^2])^{3/2} = p^{3/2}$, and hence in order to make the right-hand side of (4) to be $o(1)$, we at least need $p = o(n^{1/3})$ when $\mathcal{A}$ is the class of Euclidean balls and $p = o(n^{2/7})$ when $\mathcal{A}$ is the class of all convex sets. Similar conditions are needed in other papers cited above. It is worthwhile to mention here that, when $\mathcal{A}$ is the class of all convex sets, it was proved by [26] that $\rho_n(\mathcal{A}) \ge c\mathrm{E}[\|X_i\|^3]/\sqrt{n}$ for some universal constant $c > 0$.

In modern statistical applications, such as high dimensional estimation and multiple hypothesis testing, however, p is often larger or even much larger than n. It is therefore interesting to ask whether it is possible to provide a nontrivial class of sets $\mathcal{A}$ in $\mathbb{R}^p$ for which we would have

$$\rho_n(\mathcal{A}) \to 0 \text{ even if } p \text{ is potentially larger or much larger than } n. \tag{5}$$

In this paper, we derive bounds on $\rho_n(\mathcal{A})$ for $\mathcal{A} = \mathcal{A}^r$ being the class of all rectangles, or more generally for $\mathcal{A} = \mathcal{A}^s$ being the class of simple convex sets, and show that these bounds lead to (5). We call any convex set a simple convex set if it can be well approximated by an affine transformation of a rectangle. Once we establish the result for rectangles, we extend the results to the simple convex sets, which include an interesting and important class of sparsely convex sets. These are sets that can be represented as an intersection of many convex sets whose indicator functions depend nontrivially only on a small subset of their arguments.

The sets considered are useful for applications in mathematical statistics. In particular, the rectangles and sparsely convex sets are interesting because they allow us to approximate the probabilities of various key statistics exceeding or falling below certain thresholds. For example, the probability that a collection of Kolmogorov-type statistics falls below a collection of thresholds,

$$\mathrm{P}\left(\max_{j\in J_k} S^X_{nj} \le t_k;\ k=1,...,\kappa\right) = \mathrm{P}\left(S^X_n \in A\right),$$

can be approximated by $\mathrm{P}(S^Y_n \in A)$ within the error margin $\rho_n(\mathcal{A}^r)$; here $\{J_k\}$ are subsets of $\{1,...,p\}$, $\{t_k\}$ are the thresholds in the interval $[-\infty,\infty]$, $1 \le \kappa \le 2^p$ is an integer, and $A \subset \mathcal{A}^r$ is a rectangular region of the form $\{w \in \mathbb{R}^p : \max_{j\in J_k} w_j \le t_k;\ k=1,...,\kappa\}$. Or, for example, the probability that a collection of Pearson-type statistics falls below a collection of thresholds,

$$\mathrm{P}\Big(\|(S^X_{nj})_{j\in J_k}\|^2 \le t_k;\ k=1,...,\kappa\Big) = \mathrm{P}\left(S^X_n \in A\right)$$

can be approximated by $\mathrm{P}(S^Y_n \in A)$ within the error margin $\rho_n(\mathcal{A}^s)$; here $\{J_k\}$ are subsets of $\{1,...,p\}$ of fixed cardinality s_0, $\{t_k\}$ are the thresholds in the interval $(0,\infty]$, $1 \le \kappa \le \binom{p}{s_0}$ is an integer, and $A \subset \mathcal{A}^s$ is a sparsely convex set of the form $\{w \in \mathbb{R}^p : \|(w_j)_{j\in J_k}\|^2 \le t_k;\ k=1,...,\kappa\}$. In practice, as we demonstrate, the approximations above could be estimated using the empirical or multiplier bootstrap.

The results in this paper extend those obtained in [14] where we considered the class $\mathcal{A} = \mathcal{A}^m$ of sets of the form $A = \{w \in \mathbb{R}^p : \max_{j\in J} w_j \le a\}$ for some $a \in \mathbb{R}$ and $J \subset \{1,...,p\}$, but to obtain much better dependence on n, we employ new techniques. Most notably, we employ an induction argument as the main ingredient in the new proof, as inspired by Bolthausen [8]. Our paper builds upon our previous work [14], which in turn builds on a number of works listed in the bibliography (see [15] for a detailed review and links to the literature).

1.1. **Organization of the paper.** In Section 2, we derive a Central Limit Theorem (CLT) for rectangles in high dimensions; that is, we derive a bound on $\rho_n(\mathcal{A})$ for $\mathcal{A} = \mathcal{A}^r$ being the class of all rectangles and show that the bound converges to zero under certain conditions even when p is potentially larger or much larger than n. In Section 3, we extend this result by showing that similar bounds apply for $\mathcal{A} = \mathcal{A}^s$ being a class of simple convex sets. In Section 4, we derive high dimensional Empirical and Multiplier Bootstrap CLTs that allow to approximate the distribution of $\mathrm{P}(S^Y_n \in A)$ for $A \in \mathcal{A}^r$ or $\mathcal{A}^s$ using the data $X_1,\dots,X_n$. In Section 5, we state an induction lemma, a key result underlying all the derivations in the paper. Finally, we provide all proofs as well as some technical results in the Appendix.

1.2. **Notation.** For $a \in \mathbb{R}$, $[a]$ denotes the largest integer smaller than or equal to a. For $w = (w_1,\dots,w_p)' \in \mathbb{R}^p$ and $s = (s_1,\dots,s_p)' \in \mathbb{R}^p$, we write $w \le s$ if $w_j \le s_j$ for all $j = 1,\dots,p$. For $s = (s_1,\dots,s_p)' \in \mathbb{R}^p$ and $a \in \mathbb{R}$, we write $s + a := (s_1 + a,\dots,s_p + a)'$. Throughout the paper, $\mathbb{E}_n[\cdot]$ denotes the average over index $i = 1,\dots,n$; that is, it simply abbreviates the notation $n^{-1}\sum_{i=1}^n[\cdot]$. For example, $\mathbb{E}_n[x_{ij}] = n^{-1}\sum_{i=1}^n x_{ij}$. Also, we denote $X_1^n := (X_1,\dots,X_n)$. Finally, following standard notation, for $\alpha > 0$, we define the function $\psi_\alpha : [0,\infty) \to [0,\infty)$ by $\psi_\alpha(x) := \exp(x^\alpha) - 1$, and for a random variable ξ, we define

$$\|\xi\|_{\psi_\alpha} := \inf\{\lambda > 0 : \mathrm{E}[\psi_\alpha(|\xi|/\lambda)] \le 1\}.$$

For $\alpha \ge 1$, $\|\cdot\|_{\psi_\alpha}$ is an Orlicz norm, while for $\alpha \in (0,1)$, $\|\cdot\|_{\psi_\alpha}$ is not a norm but a quasi-norm, that is, there exists a constant K_α depending only on α such that $\|\xi_1 + \xi_2\|_{\psi_\alpha} \le K_\alpha(\|\xi_1\|_{\psi_\alpha} + \|\xi_2\|_{\psi_\alpha})$. Throughout the paper, we assume that $n \ge 4$ and $p \ge 2$.

2. High Dimensional CLT for Rectangles

This section presents a high dimensional CLT for rectangles. Let $\mathcal{A}^r$ be the class of all rectangles in $\mathbb{R}^p$; that is, $\mathcal{A}^r$ consists of all sets A of the form

$$A = \{w \in \mathbb{R}^p : a_j \le w_j \le b_j \text{ for all } j = 1,\dots,p\} \tag{6}$$

for some $-\infty \le a_j \le b_j \le \infty$, $j = 1,\dots,p$. We will derive a bound on $\rho_n(\mathcal{A}^r)$, and show that under certain conditions it leads to $\rho_n(\mathcal{A}^r) \to 0$ even when $p = p_n$ is potentially larger or much larger than n.

To describe the bound, we need to prepare some notation. Define

$$L_n := \max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^n \mathrm{E}[|X_{ij}|^3], \tag{7}$$

and for $\phi \ge 1$, define

$$M_{n,X}(\phi) := \frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\max_{1\le j\le p}|X_{ij}|^3 1\left\{\max_{1\le j\le p}|X_{ij}| > \sqrt{n}/(4\phi\log p)\right\}\right]. \tag{8}$$

Similarly, define $M_{n,Y}(\phi)$ with X_{ij}'s replaced by Y_{ij}'s in (8), and let $M_n(\phi) := M_{n,X}(\phi) + M_{n,Y}(\phi)$. The following is the first main result of this paper.

Theorem 2.1 (High Dimensional CLT for Rectangles)**.** *Suppose that there exists some constant $b > 0$ such that $m^{-1}\sum_{i=1}^m \mathrm{E}[X_{ij}^2] \ge b$ for all $j = 1,\dots,p$ and $m = [n - \log n - 1],\dots,n$. Then there exist constants $K_1, K_2 > 0$ depending only b such that for every constant $\overline{L}_n \ge L_n$,*

$$\rho_n(\mathcal{A}^r) \le K_1\left[\left(\frac{\overline{L}_n^2\log^7 p}{n}\right)^{1/6} + \frac{M_n(\phi_n)}{\overline{L}_n}\right], \tag{9}$$

with

$$\phi_n := K_2\left(\frac{\overline{L}_n^2\log^4 p}{n}\right)^{-1/6}. \tag{10}$$

Remark 2.1 (Key features of Theorem 2.1, I)**.** The bound (9) should be contrasted with Bentkus's [6] bound (4). For the sake of exposition, assume that the vectors $X_1,\dots,X_n$ are such that $\mathrm{E}[X_{ij}^2] = 1$ and for some constant

$B_n \geq 1$, $|X_{ij}| \leq B_n$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$. Then it can be shown that the bound (9) reduces to

$$\rho_n(\mathcal{A}^r) \leq K \left(\frac{B_n^2 \log^7(pn)}{n} \right)^{1/6} \tag{11}$$

for some universal constant K; see Corollary 2.1 below. Importantly, the right-hand side of (11) converges to zero even when p is much larger than n; indeed we just need $B_n^2 \log^7(pn) = o(n)$ to make $\rho_n(\mathcal{A}^r) \to 0$, and if in addition $B_n = O(1)$, the condition reduces to $\log p = o(n^{1/7})$. In contrast, Bentkus's bound (4) requires $p = o(n^{2/7})$ to make $\rho_n(\mathcal{A}) \to 0$ when $\mathcal{A}$ is the class of all convex sets. Thus, if we restrict attention to the smaller class of sets, $\mathcal{A} = \mathcal{A}^r$, the requirements on p are considerably weaker. ■

Remark 2.2 (Key features of Theorem 2.1, II)**.** On the other hand, the bound in (11) depends on n through $n^{-1/6}$, so that our Theorem 2.1 does not recover the Berry-Esseen bound when p is fixed. However, given the that the rate $n^{-1/6}$ is optimal (in a minimax sense) in CLT in infinite dimensional Banach spaces (see [4]), the factor $n^{-1/6}$ seems nearly optimal in terms of dependence on n in the high-dimensional settings. In addition, examples in [16] suggest that dependence on B_n is also optimal. Therefore, we conjecture that

$$K \left(\frac{B_n^2 \log^a(p)}{n} \right)^{1/6}$$

for some $a > 0$ is an optimal bound (in a minimax sense) in the high dimensional setting. A value of $a = 3$ could be motivated by the theory of moderate deviations for self-normalized sums when the components of X_i are independent. ■

Remark 2.3 (Relation to previous work)**.** Theorem 2.1 extends the result of Theorem 2.2 in [14] where we derived a bound on $\rho_n(\mathcal{A}^m)$ with $\mathcal{A}^m \subset \mathcal{A}^r$ consisting of all sets of the form

$$A = \{w = (w_1, \dots, w_p)' \in \mathbb{R}^p : w_j \leq a \text{ for all } j = 1, \dots, p\}$$

for some $a \in \mathbb{R}$. In particular, we improve dependence on n from $n^{-1/8}$ in [14] to $n^{-1/6}$. In addition, we note that extension to the class $\mathcal{A}^r$ from the class $\mathcal{A}^m$ is not immediate since in both papers we assume that $\mathrm{var}(S^X_{nj})$ is bounded below from zero uniformly in $j = 1, \dots, p$, so that it is not possible to obtain the class $\mathcal{A}^r$ from the class $\mathcal{A}^m$ by rescaling some coordinates of S^X_n. ■

The bound (9) depends on $M_n(\phi_n)$ whose values are problem specific. Therefore, we now apply Theorem 2.1 in two specific examples that are most useful in mathematical statistics (as well as other related fields such as econometrics). Let $b > 0$ be some constant and let $B_n \geq 1$ be a sequence of constants, possibly growing to infinity as $n \to \infty$. Assume that the following conditions are satisfied:

(M.1) $n^{-1}\sum_{i=1}^n \mathrm{E}[X_{ij}^2] \geq b$ for all $j = 1, \dots, p$,

(M.2) $n^{-1}\sum_{i=1}^n \mathrm{E}[|X_{ij}|^{2+k}] \leq B_n^k$ for all $j = 1, \dots, p$ and $k = 1, 2$.

We consider examples where one of the following conditions holds:

(E.1) $\mathrm{E}[\exp(|X_{ij}|/B_n)] \leq 2$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$,

(E.2) $\mathrm{E}[(\max_{1\leq j\leq p}|X_{ij}|/B_n)^q] \leq 2$ for all $i = 1, \dots, n$,

where $q \geq 4$. Application of Theorem 2.1 under these conditions leads to the following corollary.

Corollary 2.1 (Leading Examples)**.** *Assume that conditions* (M.1) *and* (M.2) *are satisfied. Then under* (E.1)*, we have*

$$\rho_n(\mathcal{A}^r) \leq C\left(\frac{B_n^2 \log^7(pn)}{n}\right)^{1/6}, \tag{12}$$

where the constant C depends only on b, while under (E.2)*, we have*

$$\rho_n(\mathcal{A}^r) \leq C\left[\left(\frac{B_n^2 \log^7(pn)}{n}\right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}}\right)^{1/3}\right], \tag{13}$$

where the constant C depends only on b and q.

3. High Dimensional CLT for Simple Convex Sets

In this section, we extend the results of Section 2 by considering larger classes of sets; in particular, we consider classes of simple convex sets, and obtain, under certain conditions, bounds that are similar to those in Section 2 (Corollary 3.1). Although an extension to simple convex sets is not difficult, in high dimensional spaces, the class of simple convex sets is rather large, and, as we demonstrate in Lemma 3.1, contains interesting classes of sparsely convex sets (Definition 3.1). These classes in turn may be of interest in mathematical statistics where sparse models and techniques have been getting very popular in the past years.

Consider a convex set $A \subset \mathbb{R}^p$. This set can be characterized by its support function:

$$S_A : \mathbb{S}^{p-1} \to \mathbb{R} \cup \{\infty\}, \quad v \mapsto S_A(v) := \sup\{w'v : w \in A\},$$

where $\mathbb{S}^{p-1} := \{v \in \mathbb{R}^p : \|v\| = 1\}$; in particular, $A = \cap_{v\in\mathbb{S}^{p-1}}\{w \in \mathbb{R}^p : w'v \leq S_A(v)\}$. We say that a convex set A is *m-generated* if it is generated by intersections of m half-spaces. The support function S_A of such a set A can be characterized completely by its values $\{S_A(v), v \in \mathcal{V}(A)\}$ for the set $\mathcal{V}(A)$ consisting of m unit vectors that are outward normal to the faces of A. Indeed,

$$A = \cap_{v\in\mathcal{V}(A)}\{w \in \mathbb{R}^p : w'v \leq S_A(v)\}.$$

For $\epsilon > 0$ and an m-generated convex set A^m, we define

$$A^{m,\epsilon} := \cap_{v\in\mathcal{V}(A^m)}\{w \in \mathbb{R}^p : w'v \leq S_{A^m}(v) + \epsilon\}.$$

Further, we say that a convex set A admits an approximation with precision ϵ by an m-generated convex set A^m if

$$A^m \subset A \subset A^{m,\epsilon}.$$

Let $a, d > 0$ be some constants. Let $\mathcal{A}^s$ be a class of sets A in $\mathbb{R}^p$ that satisfy the following condition:

(C) *The set A admits an approximation with precision $\epsilon = a/n$ by an m-generated convex set A^m where $m \leq (pn)^d$.*

We refer to sets A that satisfy condition (C) as simple convex sets because they can be well approximated by affine transformations of rectangles. Note that rectangles $A \in \mathcal{A}^r$ trivially satisfy condition (C) with $a = 0$ and $d = 1$.

For all $A \in \mathcal{A}^s$ with the approximating m-generated set A^m as in condition (C) and $\widetilde{X}_i = (\widetilde{X}_{i1}, \dots, \widetilde{X}_{im})' = (v'X_i)_{v \in \mathcal{V}(A^m)}$, $i = 1, \dots, n$, we assume that the following conditions are satisfied:

(M.1′) $n^{-1} \sum_{i=1}^n \mathrm{E}[\widetilde{X}_{ij}^2] \geq b$ for all $j = 1, \dots, m$,

(M.2′) $n^{-1} \sum_{i=1}^n \mathrm{E}[|\widetilde{X}_{ij}|^{2+k}] \leq B_n^k$ for all $j = 1, \dots, m$ and $k = 1, 2$.

In addition, we assume that one of the following conditions hold:

(E.1′) $\mathrm{E}[\exp(|\widetilde{X}_{ij}|/B_n)] \leq 2$ for all $i = 1, \dots, n$ and $j = 1, \dots, m$,

(E.2′) $\mathrm{E}[(\max_{1 \leq j \leq m} |\widetilde{X}_{ij}|/B_n)^q] \leq 2$ for all $i = 1, \dots, n$,

Conditions (M.1′), (M.2′), (E.1′), and (E.2′) are similar to those used in the previous section but they apply to vectors $\widetilde{X}_1, \dots, \widetilde{X}_n$ rather than to vectors $X_1, \dots, X_n$.

Recall the definition of $\rho_n(\mathcal{A})$ in (3). A simple extension of Corollary 2.1 gives the following result for the classes of sets $\mathcal{A} = \mathcal{A}^s$.

Corollary 3.1 (High dimensional CLT for simple convex sets)**.** *If all sets A in the class $\mathcal{A}^s$ satisfy* (C), (M.1′), (M.2′), *and* (E.1′), *then we have*

$$\rho_n(\mathcal{A}^s) \leq C \left(\frac{B_n^2 \log^7(pn)}{n} \right)^{1/6}, \tag{14}$$

where the constant C depends only on b, a, and d; and if all sets A in the class $\mathcal{A} = \mathcal{A}^s$ satisfy (C), (M.1′), (M.2′), *and* (E.2′), *then we have*

$$\rho_n(\mathcal{A}^s) \leq C \left[\left(\frac{B_n^2 (\log^7(pn)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}} \right)^{1/3} \right],$$

where the constant C depends only on b, q, a, and d.

Applying Corollary 3.1 to log-concave distributions gives the following particularly useful result:

Corollary 3.2 (High dimensional CLT for simple convex sets with log–concave distributions)**.** *Assume that the vectors $X_1, \dots, X_n$ have (centered) log-concave distributions on $\mathbb{R}^p$ and that maximal eigenvalue of* $\mathrm{E}[X_i X_i']$ *is bounded from above by a constant K for all $i = 1, \dots, n$. If all sets A in the class $\mathcal{A}^s$ satisfy* (C) *and* (M.1′)*, then we have*

$$\rho_n(\mathcal{A}^s) \leq C \left(\frac{\log^7(pn)}{n} \right)^{1/6},$$

where the constant C depends only on b, a, d, and K.

As we mentioned above, the class of simple convex sets is large and, in high dimensions, contains many interesting sets. In particular, this class contains sparsely convex sets defined as follows.

Definition 3.1 (Sparsely convex sets)**.** For integers $s, Q > 0$, we say that $A \subset \mathbb{R}^p$ is an (s, Q)-sparsely convex set if $A = \cap_{q=1}^{Q} A_q$ where for each q, $A_q \subset \mathbb{R}^p$ is a convex set whose indicator function $w \mapsto I(w \in A_q)$ depends at most on s components of its argument $w = (w_1, \dots, w_p)$ (which we call main components of A_q). We also say that $A = \cap_{q=1}^{Q} A_q$ is a sparse representation of A.

The simplest example fitting into this definition is a rectangle as in (6), which is a $(1, 2p)$-sparsely convex set. Another simple example is the set

$$A = \{ w \in \mathbb{R}^p : v_k' w \leq a_k, \text{ for all } k = 1, \dots, m \} \tag{15}$$

for some unit vectors $v_k \in \mathbb{S}^{p-1}$ and coefficients a_k, $k = 1, \dots, m$. If the number of non-zero components of each v_k does not exceed s, this A is an (s, m)-sparsely convex set. Another, slightly more complicated, example is the set

$$A = \{ w \in \mathbb{R}^p : a_j \leq w_j \leq b_j \text{ for all } j = 1, \dots, p \text{ and } w_1^2 + w_2^2 \leq c \} \tag{16}$$

for some coefficients $-\infty \leq a_j \leq b_j \leq \infty$, $j = 1, \dots, p$, and $0 < c \leq \infty$. This A is $(2, 2p+1)$-sparsely convex set. A more complicated example is the set

$$A = \{ w \in \mathbb{R}^p : a_j \leq w_j \leq b_j, w_k^2 + w_l^2 \leq c_{kl}, \text{ for all } j, k, l = 1, \dots, p \} \tag{17}$$

for some coefficients $-\infty \leq a_j \leq b_j \leq \infty$, $0 < c_{kl} \leq \infty$, $j, k, l = 1, \dots, p$. This A is $(2, p^2 + 2p)$-sparsely convex set. Another example, given in the introduction, is

$$A = \{ w \in \mathbb{R}^p : \| (w_j)_{j \in J_k} \|^2 \leq t_k : \ k = 1, ..., \kappa \}$$

here $\{J_k\}$ are subsets of $\{1, ..., p\}$ of fixed cardinality s_0, $\{t_k\}$ are the thresholds in the interval $(0, \infty)$, and $1 \leq \kappa \leq \binom{p}{s_0}$ is an integer. This A is (s_0, p^{s_0})-sparsely convex set. In practice the approximations above could be estimated using the empirical or multiplier bootstrap.

The following lemma shows than many sparsely convex sets are simple convex sets:

Lemma 3.1 (Sparsely convex sets are simple convex sets)**.** *Assume that A is an (s,Q)-sparsely convex set containing the origin such that $\sup_{w\in A}|w|\leq R$ and such that all sets A_q in the sparse representation $A=\cap_{q=1}^{Q}A_q$ of A satisfy $-A_q\subset\mu A_q$ for some $\mu\geq 1$. Then for any $\gamma>e/8$, there exists $\epsilon_0=\epsilon_0(\gamma)>0$ such that for any $0<\epsilon<\epsilon_0$, the set A admits an approximation with precision $R\epsilon$ by an m-generated convex set A^m where*

$$m\leq Q\Big(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\Big)^{s^2}. \tag{18}$$

Therefore, if $Q\leq(pn)^{d_0}$, $R\leq(pn)^{d_0}$, and $\mu\leq(pn)^{d_0}$ for some constant $d_0\geq 1$, then there exists an absolute integer n_0 such that the set A satisfies (C) *for all $n\geq n_0$ with $a=1$ and d depending only on s and d_0.*

Remark 3.1 (On conditions of Lemma 3.1)**.** The conditions of the lemma are sufficient to establish the bound (18) but they are not necessary. These conditions arise from the fact that our derivation relies upon the approximation result of Barvinok [3]. Similar bounds can often be established under different/weaker conditions using ad hoc arguments. For example, these assumptions can be clearly avoided in the case of the set in (15).

Remark 3.2 (On condition (M.1′) for sparsely convex sets)**.** As the proof of Lemma 3.1 reveals, a simple sufficient condition for any (s,Q)-sparsely convex set A in $\mathbb{R}^p$ to obey condition (M.1′) is that $n^{-1}\sum_{i=1}^{n}\mathrm{E}[(v'X_i)^2]\geq b$ for any $v\in\mathbb{S}^{p-1}$ with non-zero components corresponding to main components of some A_q in the sparse representation $A=\cap_{q=1}^{Q}A_q$.

We conclude this section with a lemma relating conditions (M.2′), (E.1′), and (E.2′) to conditions (M.2), (E.1), and (E.2).

Lemma 3.2 (On conditions (M.2′), (E.1′), and (E.2′) for sparsely convex sets)**.** *Assume that conditions* (M.2) *and* (E.1) *(or* (E.2)*) are satisfied with constants b and B_n. Then any (s,Q)-sparsely convex set A in $\mathbb{R}^p$ obeys conditions* (M.2′) *and* (E.1′) *(or* (E.2′)*, respectively) with constants b and s^4B_n (in place of B_n) and approximating m-generated convex set A^m constructed as in Lemma 3.1.*

4. Empirical and Multiplier Bootstrap CLTs

In the last two sections, we showed that the probabilities $\mathrm{P}(S_n^X\in A)$ can be well approximated by the Gaussian analog $\mathrm{P}(S_n^Y\in A)$ under weak conditions uniformly over rectangles $A\in\mathcal{A}^r$ or simple convex sets $A\in\mathcal{A}^s$. In practice, however, the covariance matrix of S_n^Y is typically unknown, and so approximating $\mathrm{P}(S_n^X\in A)$ by $\mathrm{P}(S_n^Y\in A)$ is infeasible. Therefore, in this section, we derive two high dimensional bootstrap CLTs. These theorems allow us to further approximate the probabilities $\mathrm{P}(S_n^Y\in A)$ (and hence $\mathrm{P}(S_n^X\in A)$) by means of the bootstrap. We consider multiplier and empirical bootstrap methods.

Our first theorem is concerned with the multiplier bootstrap. Let $e_1, \dots, e_n$ be a sequence of i.i.d. $N(0,1)$ random variables that are independent of $X_1^n = \{X_1, \dots, X_n\}$. Let $\widehat{\mu}_n^X := (\widehat{\mu}_{n1}^X, \dots, \widehat{\mu}_{np}^X)' := \mathbb{E}_n[X_i]$, and consider the normalized sum:

$$S_n^{eX} := (S_{n1}^{eX}, \dots, S_{np}^{eX})' := \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i (X_i - \widehat{\mu}_n^X).$$

We are interested in bounding

$$\rho_n^{MB}(\mathcal{A}) := \sup_{A \in \mathcal{A}} \left| \mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A) \right|$$

for $\mathcal{A} = \mathcal{A}^r$ and $\mathcal{A}^s$. To state the bound, let

$$\Sigma_n^{eX} := \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{\mu}_n^X)(X_i - \widehat{\mu}_n^X)' \text{ and } \Sigma_n^Y := \frac{1}{n} \sum_{i=1}^n \mathrm{E}[Y_i Y_i']$$

denote the covariance matrices of S_n^{eX} and S_n^Y, respectively (conditional on X_1^n in case of S_n^{eX}), with $\Sigma_{n,jk}^{eX}$ and $\Sigma_{n,jk}^Y$ denoting the (j,k)th element of Σ_n^{eX} and Σ_n^Y, respectively. Also, denote the maximum norm of the difference between Σ_n^{eX} and Σ_n^Y by Δ_n:

$$\Delta_n := \max_{1 \le j,k \le p} \left| \Sigma_{n,jk}^{eX} - \Sigma_{n,jk}^Y \right|.$$

We have the following theorem for the class of rectangles $\mathcal{A} = \mathcal{A}^r$.

Theorem 4.1 (Multiplier bootstrap CLT)**.** *Assume that condition (M.1) is satisfied. Then for any constant $\bar{\Delta}_n$, we have*

$$\rho_n^{MB}(\mathcal{A}^r) \le C \bar{\Delta}_n^{1/3} (\log p)^{2/3}$$

on the event that $\Delta_n \le \bar{\Delta}_n$ where the constant C depends only on b.

We now specialize this theorem for moment conditions as in our leading examples. For brevity of the paper, we state the corollary for the case of simple convex sets $\mathcal{A} = \mathcal{A}^s$ and note that since all rectangles is a special case of simple convex sets, the same result trivially applies for the case of rectangles $\mathcal{A} = \mathcal{A}^r$ with conditions (C), (M.1′), (M.2′), and (E.j′) replaced by (M.1), (M.2), and (E.j) for $j = 1$ or 2. We have the following corollary.

Corollary 4.1 (Leading Examples)**.** *Let $\alpha \in (0, e^{-1})$ be a constant. If all sets A in the class $\mathcal{A}^s$ satisfy* (C), (M.1′), (M.2′), *and* (E.1′), *then we have with probability at least $1 - \alpha$,*

$$\rho_n^{MB}(\mathcal{A}^s) \le C \left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6}, \tag{19}$$

where the constant C depends only on b; and if all sets A in the class $\mathcal{A}^s$ satisfy (C), (M.1′), (M.2′), *and* (E.2′), *then we have with probability at least*

$1-\alpha$,

$$\rho_n^{MB}(\mathcal{A}^s) \leq C\left[\left(\frac{B_n^2 \log^5(pn)\log^2(1/\alpha)}{n}\right)^{1/6} + \left(\frac{B_n^2 \log^3 p}{\alpha^{2/q} n^{1-2/q}}\right)^{1/3}\right], \quad (20)$$

where the constant C depends only on b and q.

Our second theorem is concerned with the empirical bootstrap. Let $X_1^*, \dots, X_n^*$ be i.i.d. draws from the empirical distribution of $X_1, \dots, X_n$. Conditional on $X_1^n = \{X_1, \dots, X_n\}$, $X_1^*, \dots, X_n^*$ are i.i.d. with mean $\widehat{\mu}_n^X = \mathbb{E}_n[X_i]$. Consider the normalized sum:

$$S_n^{X^*} := (S_{n1}^{X^*}, \dots, S_{np}^{X^*})' := \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i^* - \widehat{\mu}_n).$$

We are interested in bounding

$$\rho_n^{EB}(\mathcal{A}) := \sup_{A\in\mathcal{A}} \left| \mathrm{P}(S_n^{X^*} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)\right|$$

for $\mathcal{A} = \mathcal{A}^r$ and $\mathcal{A}^s$. To state the bound, define

$$\widehat{L}_n := \max_{1\leq j\leq p} \frac{1}{n}\sum_{i=1}^n |X_{ij} - \widehat{\mu}_{nj}^X|^3, \quad (21)$$

an empirical analog of L_n. Also, for $\phi \geq 1$, define

$$\widehat{M}_{n,X}(\phi) := \frac{1}{n}\sum_{i=1}^n \max_{1\leq j\leq p} |X_{ij} - \widehat{\mu}_{nj}^X|^3 1\left\{\max_{1\leq j\leq p}|X_{ij} - \widehat{\mu}_{nj}^X| > \sqrt{n}/(4\phi\log p)\right\},$$

$$\widehat{M}_{n,Y}(\phi) := \mathrm{E}\left[\max_{1\leq j\leq p}|S_{nj}^{eX}|^3 1\left\{\max_{1\leq j\leq p}|S_{nj}^{eX}| > \sqrt{n}/(4\phi\log p)\right\} \mid X_1^n\right]$$

empirical analogs of $M_{n,X}(\phi)$ and $M_{n,Y}(\phi)$, and let $\widehat{M}_n(\phi) := \widehat{M}_{n,X}(\phi) + \widehat{M}_{n,Y}(\phi)$. Then we have the following theorem for the class of rectangles $\mathcal{A} = \mathcal{A}^r$.

Theorem 4.2 (Empirical bootstrap CLT)**.** *For any constants b, $\overline{L}_n$, and $\overline{M}_n$, we have*

$$\rho_n^{EB}(\mathcal{A}^r) \leq \rho_n^{MB}(\mathcal{A}^r) + K_1\left[\left(\frac{\overline{L}_n^2\log^7(pn)}{n}\right)^{1/6} + \frac{\overline{M}_n}{\overline{L}_n}\right]$$

on the event that $\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj}^X)^2] \geq b$ for all $j = 1, \dots, p$, $\overline{L}_n \geq \widehat{L}_n$, and $\overline{M}_n \geq \widehat{M}_n(\phi_n)$, where

$$\phi_n := K_2\left(\frac{\overline{L}_n^2 \log^4 p}{n}\right)^{-1/6},$$

and where the constants K_1, K_2 depend only on b.

As in the case of multiplier boostrap, we now specialize this theorem for moment conditions as in our main examples. For brevity of the paper, we state the corollary for the case of simple convex sets $\mathcal{A} = \mathcal{A}^s$ and note that the same result applies for the case of rectangles as a special case. We have the following corollary.

Corollary 4.2 (Leading Examples)**.** *Let $\alpha \in (0, e^{-1})$ be a constant. Assume that $\log(1/\alpha) \leq K \log(pn)$ for some other constant K. If all sets A in the class $\mathcal{A}^s$ satisfy* (C)*,* (M.1′)*,* (M.2′)*, and* (E.1′)*, then we have with probability at least $1 - \alpha$,*

$$\rho_n^{EB}(\mathcal{A}^s) \leq C \left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6}, \tag{22}$$

where the constant C depends only on b and K; and if all sets A in the class $\mathcal{A}^s$ satisfy (C)*,* (M.1′)*,* (M.2′)*, and* (E.2′)*, then we have with probability at least $1 - \alpha$,*

$$\rho_n^{EB}(\mathcal{A}^s) \leq C \left[\left(\frac{B_n^2 \log^5(pn) \log^2(1/\alpha)}{n} \right)^{1/6} + \left(\frac{B_n^2 \log^3(pn)}{\alpha^{2/q} n^{1-2/q}} \right)^{1/3} \right], \tag{23}$$

where the constant C depends only on b, q, and K.

Remark 4.1 (The a.s. bootstrap CLTs)**.** Corollaries 4.1 and 4.2 are sharp enough to yield the following a.s. multiplier and empirical bootstrap CLTs. Let $Z, Z_1, Z_2, \dots$ be i.i.d. random variables taking values in a measurable space $(S, \mathcal{S})$. For each $n \geq 4$, let $\mathcal{F}_n$ be a class consisting of $p = p_n = |\mathcal{F}_n| \geq 2$ measurable functions $S \to \mathbb{R}$. Consider random vectors in $\mathbb{R}^p$, $X_i := X_{i,n} := (f(Z_i) - \mathrm{E}[f(Z)])_{f \in \mathcal{F}_n}$, for $i = 1, \dots, n$. Also define S_n^X, S_n^{eX}, $S_n^{X^*}$, $\rho_n^{MB}(\mathcal{A}^s)$, and $\rho_n^{EB}(\mathcal{A}^s)$ as above. Applying Corollaries 4.1 and 4.2 with $\alpha = \alpha_n = n^{-1} \log^{-2} n$ and using the Borel-Cantelli emma (recall that $\sum_{n=4}^{\infty} n^{-1} \log^{-2} n < \infty$) implies that under conditions (C), (M.1′), (M.2′), and (E.1′),

$$\rho_n^{MB}(\mathcal{A}^s) = o(1) \text{ and } \rho_n^{EB}(\mathcal{A}^s) = o(1) \text{ a.s.} \tag{24}$$

if $B_n^2 \log^7(pn)/n = o(1)$. Similarly, under conditions (C), (M.1′), (M.2′), and (E.2′), (24) holds if $B_n^2 \log^7(pn)/n = o(1)$ and $B_n^2 \log^3(pn) \log^{4/q} n / n^{1-4/q} = o(1)$.

5. Induction Lemma

In this section, we prove an induction lemma that plays a key role in the proof of our high dimensional CLT for rectangles (Theorem 2.1). Fix n, and consider normalized partial sums for $m = 1, \dots, n$:

$$S^X_{n,m} := (S^X_{n,m,1}, \dots, S^X_{n,m,p})' := \frac{1}{\sqrt{m}} \sum_{i=1}^m X_i,$$

$$S^Y_{n,m} := (S^Y_{n,m,1}, \dots, S^Y_{n,m,p})' := \frac{1}{\sqrt{m}} \sum_{i=1}^m Y_i.$$

Define

$$\varrho_{n,m} := \sup_{s \in \mathbb{R}^p, v \in [0,1]} \left| \mathrm{P}\left(\sqrt{v} S^X_{n,m} + \sqrt{1-v} S^Y_{n,m} \le s\right) - \mathrm{P}(S^Y_{n,m} \le s) \right|, \quad (25)$$

and let $M_n(\phi) := M_{n,X}(\phi) + M_{n,Y}(\phi)$. Our induction lemma below provides a bound on $\varrho_{n,m}$ in terms of $\varrho_{n,m-1}$ for each $m = [n - \log n], \dots, n$.

Lemma 5.1 (Induction Lemma)**.** *Suppose that there exists some constant $b > 0$ such that $m^{-1} \sum_{i=1}^m \mathrm{E}[X_{ij}^2] \ge b$ for all $j = 1, \dots, p$ and $m = [n - \log n - 1], \dots, n$. Then $\varrho_{n,m}$ satisfies the following inequality for all $\phi \ge 1$ and $m = [n - \log n], \dots, n$:*

$$\varrho_{n,m} \lesssim \frac{\phi^2 (\log p)^2}{n^{1/2}} \Big(L_n \phi \varrho_{n,m-1} + L_n (\log p)^{1/2} + \phi M_n(\phi) \Big) + \frac{(\log p)^{1/2}}{\phi}$$

up to a constant K that depends only on b.

Lemma 5.1 has an immediate corollary. Indeed, define

$$\varrho'_{n,m} := \sup_{A \in \mathcal{A}^r, v \in [0,1]} \left| \mathrm{P}(\sqrt{v} S^X_{n,m} + \sqrt{1-v} S^Y_{n,m} \in A) - \mathrm{P}(S^Y_{n,m} \in A) \right|$$

where $\mathcal{A}^r$ is the class of all rectangles in $\mathbb{R}^p$. Then we have

Corollary 5.1. *Suppose that there exists some constant $b > 0$ such that $m^{-1} \sum_{i=1}^m \mathrm{E}[X_{ij}^2] \ge b$ for all $j = 1, \dots, p$ and $m = [n - \log n - 1], \dots, n$. Then $\varrho'_{n,m}$ satisfies the following inequality for all $\phi \ge 1$ and $m = [n - \log n], \dots, n$:*

$$\varrho'_{n,m} \lesssim \frac{\phi^2 (\log p)^2}{n^{1/2}} \Big(L_n \phi \varrho'_{n,m-1} + L_n (\log p)^{1/2} + \phi M_n(2\phi) \Big) + \frac{(\log p)^{1/2}}{\phi}$$

up to a constant K' that depends only on b.

Appendix A. Technical tools

A.1. Anti-concentration inequality.

Lemma A.1 (Nazarov inequality, [27])**.** *Let $Y = (Y_1, \dots, Y_p)'$ be a centered Gaussian random vector in $\mathbb{R}^p$ such that $\mathrm{E}[Y_j^2] \ge b$ for all $j = 1, \dots, p$ and some constant $b > 0$. Then for every $s \in \mathbb{R}^p$ and $a > 0$,*

$$\mathrm{P}(Y \le s + a) - \mathrm{P}(Y \le s) \le C a (\log p)^{1/2},$$

where C is a constant depending only on b.

Remark A.1. This inequality is less sharp than the dimension-free anti-concentration bound $Ca\mathrm{E}\|Y\|_\infty$ proved in [17] for the case of max rectangles (anti-concentration inequalities for suprema of Gaussian processes). However, the former inequality allows for more general rectangles than the latter. The difference in sharpness for the case of max-rectangles arises due to dimension-dependence $(\log p)^{1/2}$, in particular the term $(\log p)^{1/2}$ can be much larger than $\mathrm{E}\|Y\|_\infty$. This also makes the anti-concentration bound in [17] more relevant for the study of suprema of Gaussian processes indexed by infinite classes. It is an interesting question whether one could establish a dimension-free anti-concentration bound similar to that in [17] for classes of rectangular sets other than max rectangles. ∎

Proof of Lemma A.1. Let $\Sigma = \mathrm{E}[YY']$, so that Y has the same distribution as $\Sigma^{1/2}Z$ where Z is a standard Gaussian random vector. Denote by σ_j the jth column of $\Sigma^{1/2}$, so that $\Sigma^{1/2} = (\sigma_1, \dots, \sigma_p)$. Then

$$\begin{aligned} \mathrm{P}(Y \le s + a) &= \mathrm{P}(\Sigma^{1/2}Z \le s + a) \\ &= \mathrm{P}((\sigma_j/|\sigma_j|)'Z \le (s_j + a)/|\sigma_j| \text{ for all } j = 1, \dots, p), \end{aligned}$$

and

$$\mathrm{P}(Y \le s) = \mathrm{P}((\sigma_j/|\sigma_j|)'Z \le s_j/|\sigma_j| \text{ for all } j = 1, \dots, p).$$

Since Z is a standard Gaussian random vector, and $a/\|\sigma_j\| \le a/b^{1/2}$ for all $j = 1, \dots, p$, the assertion follows from Theorem 20 in [22], whose proof the authors credit to Nazarov [27]. ∎

A.2. **Maximal inequalities.**

Lemma A.2. *Let $X_1, \dots, X_n$ be independent centered random vectors in $\mathbb{R}^p$ with $p \ge 2$. Define $Z := \max_{1\le j\le p} |\sum_{i=1}^n X_{ij}|$, $M := \max_{1\le i\le n} \max_{1\le j\le p} |X_{ij}|$ and $\sigma^2 := \max_{1\le j\le p} \sum_{i=1}^n \mathrm{E}[X_{ij}^2]$. Then*

$$\mathrm{E}[Z] \le K(\sigma\sqrt{\log p} + \sqrt{\mathrm{E}[M^2]}\log p).$$

where K is a universal constant.

Proof. See Lemma 8 in [17]. ∎

Lemma A.3. *Assume the setting of Lemma A.2. (i) For every $\eta > 0, \beta \in (0, 1]$ and $t > 0$,*

$$\mathrm{P}\{Z \ge (1+\eta)\mathrm{E}[Z] + t\} \le \exp\{-t^2/(3\sigma^2)\} + 3\exp\{-(t/(K\|M\|_{\psi_\beta}))^\beta\},$$

where $K = K(\eta, \beta)$ is a constant depending only on η, β.

(ii) For every $\eta > 0, s \ge 1$ and $t > 0$,

$$\mathrm{P}\{Z \ge (1+\eta)\mathrm{E}[Z] + t\} \le \exp\{-t^2/(3\sigma^2)\} + K'\mathrm{E}[M^s]/t^s,$$

where $K' = K'(\eta, s)$ is a constant depending only on η, s.

Proof. See Theorem 4 in [1] for case (i) and Theorem 2 in [2] for case (ii). See also [19]. ∎

Lemma A.4. *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$ such that $X_{ij} \geq 0$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$. Define $Z := \max_{1 \leq j \leq p} \sum_{i=1}^n X_{ij}$ and $M := \max_{1 \leq i \leq n} \max_{1 \leq j \leq p} X_{ij}$. Then*

$$\mathrm{E}[Z] \leq K \left(\max_{1 \leq j \leq p} \mathrm{E}[\textstyle\sum_{i=1}^n X_{ij}] + \mathrm{E}[M] \log p \right),$$

where K is a universal constant.

Proof. See Lemma 9 in [17]. ∎

Lemma A.5. *Assume the setting of Lemma A.4. (i) For every $\eta > 0, \beta \in (0,1]$ and $t > 0$,*

$$\mathrm{P}\{Z \geq (1+\eta)\mathrm{E}[Z] + t\} \leq 3 \exp\{-(t/(K\|M\|_{\psi_\beta}))^\beta\},$$

where $K = K(\eta, \beta)$ is a constant depending only on η, β. (ii) For every $\eta > 0, s \geq 1$ and $t > 0$,

$$\mathrm{P}\{Z \geq (1+\eta)\mathrm{E}[Z] + t\} \leq K' \mathrm{E}[M^s]/t^s,$$

where $K' = K'(\eta, s)$ is a constant depending only on η, s.

The proof of Lemma A.5 relies on the following lemma, which follows from Theorem 10 in [24].

Lemma A.6. *Assume the setting of Lemma A.4. Suppose that there exists a constant B such that $M \leq B$. Then for every $\eta, t > 0$,*

$$\mathrm{P}\left\{ Z \geq (1+\eta)\mathrm{E}[Z] + B\left(\frac{2}{3} + \frac{1}{\eta}\right) t \right\} \leq e^{-t}.$$

Proof of Lemma A.6. By homogeneity, we may assume that $B = 1$. Then by Theorem 10 in [24], for every $\lambda > 0$,

$$\log \mathrm{E}[\exp(\lambda(Z - \mathrm{E}[Z]))] \leq \varphi(\lambda)\mathrm{E}[Z],$$

where $\varphi(\lambda) = e^\lambda - \lambda - 1$. Hence by Markov's inequality, with $a = \mathrm{E}[Z]$,

$$\mathrm{P}\{Z - \mathrm{E}[Z] \geq t\} \leq e^{-\lambda t + a\varphi(\lambda)}.$$

The right side is minimized at $\lambda = \log(1 + t/a)$, at which $-\lambda t + a\varphi(\lambda) = -aq(t/a)$ where $q(t) = (1+t)\log(1+t) - t$. It is routine to verify that $q(t) \geq t^2/(2(1+t/3))$, so that

$$\mathrm{P}\{Z - \mathrm{E}[Z] \geq t\} \leq e^{-\frac{t^2}{2(a+t/3)}}.$$

Solving $t^2/(2(a+t/3)) = s$ gives $t = s/3 + \sqrt{s^2/9 + 2as} \leq 2s/3 + \sqrt{2as}$. Therefore, we have

$$\mathrm{P}\{Z \geq \mathrm{E}[Z] + \sqrt{2as} + 2s/3\} \leq e^{-s}.$$

The conclusion follows from the inequality $\sqrt{2as} \leq \eta a + \eta^{-1} s$. ∎

Proof of Lemma A.5. The proof is a modification of that of Theorem 4 in [1] (or Theorem 2 in [2]). We begin with noting that we may assume that $(1+\eta)8\mathrm{E}[M] \leq t/4$, since otherwise we can make the lemma trivial by setting K or K' large enough. Take

$$\rho = 8\mathrm{E}[M],\ Y_{ij} = \begin{cases} X_{ij}, & \text{if } \max_{1 \leq j \leq p} X_{ij} \leq \rho, \\ 0, & \text{otherwise} \end{cases}$$

Define

$$W_1 = \max_{1 \leq j \leq p} \sum_{i=1}^n Y_{ij},\ W_2 = \max_{1 \leq j \leq p} \sum_{i=1}^n (X_{ij} - Y_{ij}).$$

Then

$$\begin{aligned} \mathrm{P}\{Z \geq (1+\eta)\mathrm{E}[Z] + t\} &\leq \mathrm{P}\{W_1 \geq (1+\eta)\mathrm{E}[Z] + 3t/4\} + \mathrm{P}(W_2 \geq t/4) \\ &\leq \mathrm{P}\{W_1 \geq (1+\eta)\mathrm{E}[W_1] - (1+\eta)\mathrm{E}[W_2] + 3t/4\} + \mathrm{P}(W_2 \geq t/4). \end{aligned}$$

Observe that

$$\mathrm{P}\left\{ \max_{1 \leq m \leq n} \max_{1 \leq j \leq p} \sum_{i=1}^m (X_{ij} - Y_{ij}) > 0 \right\} \leq \mathrm{P}(M > \rho) \leq 1/8,$$

so that by the Hoffmann-Jørgensen inequality [see 23, Proposition 6.8], we have

$$\mathrm{E}[W_2] \leq 8\mathrm{E}[M] \leq t/(4(1+\eta)).$$

Hence

$$\mathrm{P}\{Z \geq (1+\eta)\mathrm{E}[Z] + t\} \leq \mathrm{P}\{W_1 \geq (1+\eta)\mathrm{E}[W_1] + t/2\} + \mathrm{P}(W_2 \geq t/4).$$

By Lemma A.6, the first term on the right-hand side is bounded by $e^{-ct/\rho}$ where c depends only on η. We bound the second term separately in cases (i) and (ii). Below $C_1, C_2, \dots$ are constants that depend only on η, β, s.

Case (i). By Theorem 6.21 in [23] (note that a version of their theorem applies to nonnegative random vectors) and the fact that $\mathrm{E}[W_2] \leq 8\mathrm{E}[M]$,

$$\|W_2\|_{\psi_\beta} \leq C_1(\mathrm{E}[W_2] + \|M\|_{\psi_\beta}) \leq C_2 \|M\|_{\psi_\beta},$$

which implies that $\mathrm{P}(W_2 \geq t/4) \leq 2\exp\{-(t/(C_3\|M\|_{\psi_\beta}))^\beta\}$. Since $\rho \leq C_4\|M\|_{\psi_\beta}$, we conclude that

$$e^{-ct/\rho} + \mathrm{P}(W_2 \geq t/4) \leq 3\exp\{-(t/(C_5\|M\|_{\psi_\beta}))^\beta\}.$$

Case (ii). By Theorem 6.20 in [23] (note that a version of their theorem applies to nonnegative random vectors) and the fact that $\mathrm{E}[W_2] \leq 8\mathrm{E}[M]$,

$$(\mathrm{E}[W_2^s])^{1/s} \leq C_6(\mathrm{E}[W_2] + (\mathrm{E}[M^s])^{1/s}) \leq C_7(\mathrm{E}[M^s])^{1/s}.$$

The conclusion follows from Markov's inequality together with the simple fact that $e^{-t}/t^{-s} \to 0$ as $t \to \infty$. ∎

A.3. **Other useful inequalities.**

Lemma A.7. *Let $\varphi : \mathbb{R} \to \mathbb{R}_+$ and $\phi : \mathbb{R} \to \mathbb{R}_+$ be nondecreasing functions, and let ξ_1 and ξ_2 be independent random variables. Then*

$$\mathrm{E}[\varphi(\xi_1)]\mathrm{E}[\phi(\xi_1)] \leq \mathrm{E}[\varphi(\xi_1)\phi(\xi_1)], \tag{26}$$

$$\mathrm{E}[\varphi(\xi_1)]\mathrm{E}[\phi(\xi_2)] \leq \mathrm{E}[\varphi(\xi_1)\phi(\xi_1)] + \mathrm{E}[\varphi(\xi_2)\phi(\xi_2)], \tag{27}$$

$$\mathrm{E}[\varphi(\xi_1)\phi(\xi_2)] \leq \mathrm{E}[\varphi(\xi_1)\phi(\xi_1)] + \mathrm{E}[\varphi(\xi_2)\phi(\xi_2)]. \tag{28}$$

Moreover, (28) holds without independence of ξ_1 and ξ_2.

Proof. Inequality (26) is Chebyshev's association inequality; see Theorem 2.14 in [10]. Further, since ξ_1 and ξ_2 are independent, $\mathrm{E}[\varphi(\xi_1)\phi(\xi_2)] = \mathrm{E}[\varphi(\xi_1)]\mathrm{E}[\phi(\xi_2)]$, and so (27) follows from (28). In turn, (28) follows from

$$\begin{aligned}\mathrm{E}[\varphi(\xi_1)\phi(\xi_2)] &\leq \mathrm{E}[\varphi(\xi_1)\phi(\xi_2)] + \mathrm{E}[\phi(\xi_1)\varphi(\xi_2)] \\ &\leq \mathrm{E}[\varphi(\xi_1)\phi(\xi_1)] + \mathrm{E}[\varphi(\xi_2)\phi(\xi_2)]\end{aligned}$$

where the first inequality holds because $\phi(\xi_1)\varphi(\xi_2) \geq 0$ and the second inequality follows from rearranging terms in the following inequality:

$$\mathrm{E}[(\varphi(\xi_1) - \varphi(\xi_2))(\phi(\xi_1) - \phi(\xi_2))] \geq 0,$$

which holds by monotonicity of φ and ϕ. This completes the proof of the lemma. ∎

Lemma A.8. *Let ξ be a nonnegative random variable such that $\mathrm{P}(\xi > x) \leq Ae^{-x/B}$ for all $x \geq 0$ and for some constants $A, B > 0$. Then for every $t \geq 0$, $\mathrm{E}[\xi^3 1\{\xi > t\}] \leq 6A(t+B)^3 e^{-t/B}$.*

Proof. Observe that

$$\begin{aligned}\mathrm{E}[\xi^3 1\{\xi > t\}] &= 3\int_0^t \mathrm{P}(\xi > t)x^2 dx + 3\int_t^\infty \mathrm{P}(\xi > x)x^2 dx \\ &= \mathrm{P}(\xi > t)t^3 + 3\int_t^\infty \mathrm{P}(\xi > x)x^2 dx.\end{aligned}$$

Since $\mathrm{P}(\xi > x) \leq Ae^{-x/B}$, using integration by parts, we have

$$\int_t^\infty \mathrm{P}(\xi > s)x^2 dx \leq A(Bt^2 + 2B^2 t + 2B^3)e^{-t/B},$$

which leads to

$$\mathrm{E}[\xi^3 1\{\xi > t\}] \leq A(t^3 + 3Bt^2 + 6B^2 t + 6B^3)e^{-t/B} \leq 6A(t+B)^3 e^{-t/B}.$$

This completes the proof of the lemma. ∎

Appendix B. Proofs for Section 5

Proof of Lemma 5.1. The proof of this lemma relies on a Slepian-Stein method developed in [14] and an induction idea of [8]. In the proof, inequalities $a_n \lesssim d_n$ are understood as $a_n \leq C d_n$ where C is a constant that depends only on b. Also, since $n \geq 4$, it follows that $[n - \log n] \geq n/2$, so that $m \geq n/2$ for all $m = [n - \log n], \dots, n$. We will use this inequality in the proof frequently without additional notice.

Fix $s = (s_1, \dots, s_p)' \in \mathbb{R}^p$, $v \in [0, 1]$, and $m = [n - \log n], \dots, n$. Let $W_1, \dots, W_n$ be a copy of $Y_1, \dots, Y_n$. Without loss of generality, we assume that sequences $X_1, \dots, X_n$, $Y_1, \dots, Y_n$, and $W_1, \dots, W_n$ are independent. Consider

$$S^W_{n,m} := \frac{1}{\sqrt{m}} \sum_{i=1}^m W_i.$$

Then $\mathrm{P}(S^Y_{n,m} \leq s) = \mathrm{P}(S^W_{n,m} \leq s)$, so that

$$\varrho_{n,m} = \sup_{s \in \mathbb{R}^p, v \in [0,1]} \left| \mathrm{P}\left(\sqrt{v} S^X_{n,m} + \sqrt{1-v} S^Y_{n,m} \leq s\right) - \mathrm{P}(S^W_{n,m} \leq s) \right|.$$

Further, denote

$$\beta := \phi \log p,$$

and for $w \in \mathbb{R}^p$, define

$$F_\beta(w) := \frac{1}{\beta} \log\left(\sum_{j=1}^p \exp(\beta(w_j - s_j)) \right).$$

It is easy to check that the function $F_\beta(w)$ has the following property:

$$0 \leq F_\beta(w) - \max_{1 \leq j \leq p} (w_j - s_j) \leq \beta^{-1} \log p = \phi^{-1} \tag{29}$$

for all $w \in \mathbb{R}^p$. Also, consider a function $g_0 : \mathbb{R} \to [0, 1]$ with bounded derivatives up to the third order such that $g_0(t) = 1$ for $t \leq 0$ and $g_0(t) = 0$ for $t \geq 1$. For $t \in \mathbb{R}$, define $g(t) := g_0(\phi t)$, and for $w \in \mathbb{R}^p$, define

$$m(w) := g(F_\beta(w)).$$

For brevity of notation, we will use indices to denote partial derivatives of m; for example, $\partial_j \partial_k \partial_l m = m_{jkl}$. The function $m(w)$ has the following property established in Lemmas A.5 and A.6 of [14]: for every $j, k, l = 1, \dots, p$, there exists a function $U_{jkl}(w)$ such that

$$|m_{jkl}(w)| \leq U_{jkl}(w), \tag{30}$$

$$\sum_{j,k,l=1}^p U_{jkl}(w) \lesssim (\phi^3 + \phi\beta + \phi\beta^2) \lesssim \phi\beta^2, \tag{31}$$

$$U_{jkl}(w) \lesssim U_{jkl}(w + \widetilde{w}) \lesssim U_{jkl}(w) \tag{32}$$

where inequalities (30) and (31) hold for all $w \in \mathbb{R}^p$, and inequality (32) holds for all $w, \widetilde{w} \in \mathbb{R}^p$ with $\max_{1 \leq j \leq p} |\widetilde{w}_j| \beta \leq 1$ (formally, [14] only considered the case $s = (0, \dots, 0)'$ but the extension to $s \in \mathbb{R}^p$ is trivial). Moreover, for $w \in \mathbb{R}^p$ and $t > 0$, define

$$h(w,t) := 1\left\{-\phi^{-1} - t/\beta < \max_{1\le j\le p}(w_j - s_j) \le \phi^{-1} + t/\beta\right\}. \tag{33}$$

Finally, for $t \in (0,1)$, define

$$\omega(t) := \frac{1}{\sqrt{t} \wedge \sqrt{1-t}}.$$

The proof consists of two steps. In the first step, we show that

$$\mathrm{E}[\mathcal{I}_n] \lesssim \frac{\phi^2(\log p)^2}{n^{1/2}}\Big(L_n\phi\varrho_{n,m-1} + L_n(\log p)^{1/2} + \phi M_n(\phi)\Big) + \frac{(\log p)^{1/2}}{\phi}$$

where

$$\mathcal{I}_n := m(\sqrt{v}S_{n,m}^X + \sqrt{1-v}S_{n,m}^Y) - m(S_{n,m}^W).$$

In the second step, we combine the bound from the first step with Lemma A.1 to complete the proof.

Step 1. For $t \in [0,1]$, define the Slepian interpolant

$$Z(t) := \sum_{i=1}^m Z_i(t)$$

where

$$Z_i(t) := \frac{1}{\sqrt{m}}\left(\sqrt{t}(\sqrt{v}X_i + \sqrt{1-v}Y_i) + \sqrt{1-t}W_i\right).$$

Note that $Z(1) = \sqrt{v}S_{n,m}^X + \sqrt{1-v}S_{n,m}^Y$ and $Z(0) = S_{n,m}^W$, and so

$$\mathcal{I}_n = m(\sqrt{v}S_{n,m}^X + \sqrt{1-v}S_{n,m}^Y) - m(S_{n,m}^W) = \int_0^1 \frac{dm(Z(t))}{dt}dt. \tag{34}$$

Also, define Stein leave-one-out terms

$$Z^{(i)}(t) := Z(t) - Z_i(t).$$

Finally, define

$$\dot{Z}_i(t) := \frac{1}{\sqrt{m}}\left(\frac{1}{\sqrt{t}}(\sqrt{v}X_i + \sqrt{1-v}Y_i) - \frac{1}{\sqrt{1-t}}W_i\right).$$

For brevity of notation, we omit argument t; that is, we write $Z = Z(t)$, $Z_i = Z_i(t)$, $Z^{(i)} = Z^{(i)}(t)$, and $\dot{Z}_i = \dot{Z}_i(t)$.

Now, it follows from (34) and Taylor's theorem that

$$\mathrm{E}[\mathcal{I}_n] = \frac{1}{2}\sum_{j=1}^p\sum_{i=1}^n\int_0^1 \mathrm{E}[m_j(Z)\dot{Z}_{ij}]dt = \frac{1}{2}(I + II + III)$$

where

$$I := \sum_{j=1}^{p}\sum_{i=1}^{m}\int_0^1 \mathrm{E}[m_j(Z^{(i)})\dot{Z}_{ij}]dt,$$

$$II := \sum_{j,k=1}^{p}\sum_{i=1}^{m}\int_0^1 \mathrm{E}[m_{jk}(Z^{(i)})\dot{Z}_{ij}Z_{ik}]dt,$$

$$III := \sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1 (1-\tau)\mathrm{E}[m_{jkl}(Z^{(i)}+\tau Z_i)\dot{Z}_{ij}Z_{ik}Z_{il}]d\tau dt.$$

By independence of $Z^{(i)}$ from $\dot{Z}_{ij}$ together with $\mathrm{E}[\dot{Z}_{ij}]=0$, we have $I=0$. Also, by independence of $Z^{(i)}$ from $\dot{Z}_{ij}Z_{ik}$ together with

$$\begin{aligned}\mathrm{E}[\dot{Z}_{ij}Z_{ik}] &= \frac{1}{m}\mathrm{E}\left[(\sqrt{v}X_{ij}+\sqrt{1-v}Y_{ij})(\sqrt{v}X_{ik}+\sqrt{1-v}Y_{ik})-W_{ij}W_{ik}\right]\\ &= \frac{1}{m}\mathrm{E}[vX_{ij}X_{ik}+(1-v)Y_{ij}Y_{ik}-W_{ij}W_{ik}]=0,\end{aligned}$$

we have $II=0$. Therefore, it suffices to bound III.

To this end, denote

$$\chi_i = 1\left\{\max_{1\le j\le p}|X_{ij}|\vee|Y_{ij}|\vee|W_{ij}|\le\sqrt{n}/(4\beta)\right\}.$$

Then $III=III_1+III_2$ where

$$III_1 := \sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1 (1-\tau)\mathrm{E}[\chi_i m_{jkl}(Z^{(i)}+\tau Z_i)\dot{Z}_{ij}Z_{ik}Z_{il}]d\tau dt,$$

$$III_2 := \sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1 (1-\tau)\mathrm{E}[(1-\chi_i) m_{jkl}(Z^{(i)}+\tau Z_i)\dot{Z}_{ij}Z_{ik}Z_{il}]d\tau dt.$$

We bound III_1 and III_2 separately. For III_2, we have

$$\begin{aligned}|III_2| &\le \sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1 \mathrm{E}[(1-\chi_i)U_{jkl}(Z^{(i)}+\tau Z_i)|\dot{Z}_{ij}Z_{ik}Z_{il}|]d\tau dt\\ &\lesssim \phi\beta^2\sum_{i=1}^{m}\int_0^1\mathrm{E}[(1-\chi_i)\max_{1\le j,k,l\le p}|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\ &\lesssim \frac{\phi\beta^2}{m^{3/2}}\sum_{i=1}^{m}\int_0^1\omega(t)\mathrm{E}[(1-\chi_i)\max_{1\le j\le p}|X_{ij}|^3\vee|Y_{ij}|^3\vee|W_{ij}|^3]dt. \quad (35)\end{aligned}$$

where the first and the second lines follow from (30) and (31), respectively. Further, denoting $\mathcal{T}=\sqrt{n}/(4\beta)$, the union bound gives

$$1-\chi_i \le 1\left\{\max_{1\le j\le p}|X_{ij}|>\mathcal{T}\right\}+1\left\{\max_{1\le j\le p}|Y_{ij}|>\mathcal{T}\right\}+1\left\{\max_{1\le j\le p}|W_{ij}|>\mathcal{T}\right\}.$$

Therefore, using inequality

$$\max_{1\leq j\leq p}|X_{ij}|^3\vee|Y_{ij}|^3\vee|W_{ij}|^3\leq\max_{1\leq j\leq p}|X_{ij}|^3+\max_{1\leq j\leq p}|Y_{ij}|^3+\max_{1\leq j\leq p}|W_{ij}|^3$$

and inequality (28) of Lemma A.7, we obtain that the integral in (35) is bounded from above up to an absolute constant by

$$\mathrm{E}\left[\max_{1\leq j\leq p}|X_{ij}|^3 1\left\{\max_{1\leq j\leq p}|X_{ij}|>\mathcal{T}\right\}\right]+\mathrm{E}\left[\max_{1\leq j\leq p}|Y_{ij}|^3 1\left\{\max_{1\leq j\leq p}|Y_{ij}|>\mathcal{T}\right\}\right]$$

since W_i's have the same distribution as that of Y_i's. Conclude that

$$|III_2|\lesssim(M_{n,X}(\phi)+M_{n,Y}(\phi))\phi\beta^2/n^{1/2}=M_n(\phi)\phi\beta^2/n^{1/2}$$

since $m\geq n/2$, and so $1/m^{3/2}\leq(2/n)^{3/2}$.

To bound III_1, recall the definition of $h(w,t)$ in (33). Note that $m_{jkl}(Z^{(i)}+\tau Z_i)=0$ for all $\tau\in[0,1]$ if both $h(Z^{(i)},2)=0$ and $\chi_i=1$ hold (indeed, if $\chi_i=1$, then $\max_{1\leq j\leq p}|Z_{ij}|\leq(3/4)(n/m)^{1/2}/\beta\leq 2/\beta$, and so when both $h(Z^{(i)},2)=0$ and $\chi_i=1$ hold, we have that $h(Z^{(i)}+\tau Z_i,0)=0$, which in turn means that either $F_\beta(Z^{(i)}+\tau Z_i)\leq 0$ or $F_\beta(Z^{(i)}+\tau Z_i)\geq\phi^{-1}$ because of (29); in both cases, $g'(F_\beta(Z^{(i)}+\tau Z_i))=0$, so that the claim follows from the definition of m). Therefore,

$$\begin{aligned}|III_1|&\leq\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1\mathrm{E}[\chi_i|m_{jkl}(Z^{(i)}+\tau Z_i)\dot{Z}_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1\mathrm{E}[\chi_i h(Z^{(i)},2)U_{jkl}(Z^{(i)}+\tau Z_i)|\dot{Z}_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\int_0^1\mathrm{E}[\chi_i h(Z^{(i)},2)U_{jkl}(Z^{(i)})|\dot{Z}_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\mathrm{E}[h(Z^{(i)},2)U_{jkl}(Z^{(i)})]\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\end{aligned}\tag{36}$$

where the second inequality follows from (30), the third inequality from (32), and the fourth inequality from the indepence of $Z^{(i)}$ from $\dot{Z}_{ij}Z_{ik}Z_{il}$. Further, we split the integral in (36) inserting $\chi_i+(1-\chi_i)$ under the first expectation sign. We have

$$\begin{aligned}&\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1\mathrm{E}[(1-\chi_i)h(Z^{(i)},2)U_{jkl}(Z^{(i)})]\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\&\qquad\lesssim\phi\beta^2\sum_{i=1}^{m}\int_0^1\mathrm{E}[1-\chi_i]\mathrm{E}\left[\max_{1\leq j,k,l\leq p}|\dot{Z}_{ij}Z_{ik}Z_{il}|\right]dt\lesssim M_n(\phi)\phi\beta^2/n^{1/2}\end{aligned}$$

where the last inequality follows from an argument similar to that used to bound III_2 with an application of (26) and (27) instead of (28) in Lemma

A.7. Also, since $h(Z^{(i)},2)=0$ if both $h(Z,4)=0$ and $\chi_i=1$ hold by the same argument as above, we have

$$\begin{aligned}
&\sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1 \mathrm{E}[\chi_i h(Z^{(i)},2)U_{jkl}(Z^{(i)})]\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\
&\qquad\lesssim \sum_{j,k,l=1}^{p}\sum_{i=1}^{m}\int_0^1 \mathrm{E}[h(Z,4)U_{jkl}(Z)]\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\
&\qquad= \sum_{j,k,l=1}^{p}\int_0^1 \mathrm{E}[h(Z,4)U_{jkl}(Z)]\sum_{i=1}^{m}\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\
&\qquad\lesssim \phi\beta^2\int_0^1 \mathrm{E}[h(Z,4)]\max_{1\leq j,k,l\leq p}\sum_{i=1}^{m}\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt \qquad (37)
\end{aligned}$$

We split the integral in (37) inserting $\chi_m+(1-\chi_m)$ under the first expectation sign again. We have

$$\begin{aligned}
&\phi\beta^2\int_0^1 \mathrm{E}[(1-\chi_m)h(Z,4)]\max_{1\leq j,k,l\leq p}\sum_{i=1}^{m}\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt \qquad (38)\\
&\qquad\lesssim \phi\beta^2\sum_{i=1}^{m}\int_0^1 \mathrm{E}[1-\chi_m]\mathrm{E}\left[\max_{1\leq j,k,l\leq p}|\dot{Z}_{ij}Z_{ik}Z_{il}|\right]dt. \qquad (39)
\end{aligned}$$

Under the integral in (39), we add the term

$$\mathrm{E}[1-\chi_i]\mathrm{E}\left[\max_{1\leq j,k,l\leq p}|\dot{Z}_{mj}Z_{mk}Z_{ml}|\right],$$

and then the resulting expression can be bounded using an argument similar to that used to bound III_2 with an application of (26) and (27) instead of (28) in Lemma A.7, so that we obtain (39) $\lesssim M_n(\phi)\phi\beta^2/n^{1/2}$.

In addition, since $h(Z,4)=0$ if both $h(Z^{(m)},6)=0$ and $\chi_m=1$ hold, by the same argument as above, we have

$$\begin{aligned}
&\phi\beta^2\int_0^1 \mathrm{E}[\chi_m h(Z,4)]\max_{1\leq j,k,l\leq p}\sum_{i=1}^{m}\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\
&\quad\lesssim \phi\beta^2\int_0^1 \mathrm{E}[h(Z^{(m)},6)]\max_{1\leq j,k,l\leq p}\sum_{i=1}^{m}\mathrm{E}[|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt\\
&\quad\lesssim (\widetilde{L}_n\phi\beta^2/n^{1/2})\int_0^1 \omega(t)\mathrm{E}[h(Z^{(m)},6)]dt \lesssim \frac{L_n\phi\beta^2}{n^{1/2}}\int_0^1 \omega(t)\mathrm{E}[h(Z^{(m)},6)]dt
\end{aligned}$$

where

$$\widetilde{L}_n=\max_{1\leq j\leq p}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[|X_{ij}|^3+|Y_{ij}|^3\right]$$

and the last line follows from $m\geq n/2$ and the observation that $\mathrm{E}[|Y_{ij}|^3]\lesssim(\mathrm{E}[|Y_{ij}|^2])^{3/2}=(\mathrm{E}[|X_{ij}|^2])^{3/2}\leq\mathrm{E}[|X_{ij}|^3]$ by Gaussianity of Y_{ij}.

It remains to bound $\int_0^1 \omega(t)\mathrm{E}[h(Z^{(m)}, 6)]dt$. Note that

$$Z^{(m)} =_d \frac{1}{\sqrt{m}}\sum_{i=1}^{m-1}(\widetilde{v}^{1/2}X_i + (1-\widetilde{v})^{1/2}Y_i)$$
$$= \sqrt{\frac{m-1}{m}}(\widetilde{v}^{1/2}S^X_{n,m-1} + (1-\widetilde{v})^{1/2}S^Y_{n,m-1}) = \sqrt{\frac{m-1}{m}}\widetilde{V}_{n,m-1}$$

where $\widetilde{v} := \sqrt{tv}$ and $\widetilde{V}_{n,m-1} := \widetilde{v}^{1/2}S^X_{n,m-1} + (1-\widetilde{v})^{1/2}S^Y_{n,m-1}$. Therefore,

$$\mathrm{E}[h(Z^{(m)}, 6)] = \mathrm{P}\left(\widetilde{V}_{n,m-1} \le \bar{I}\right) - \mathrm{P}\left(\widetilde{V}_{n,m-1} \le \underline{I}\right)$$

where $\underline{I}$ and $\bar{I}$ are vectors in $\mathbb{R}^p$ defined as

$$\underline{I} := \sqrt{\frac{m}{m-1}}(s - \phi^{-1} - 6\beta^{-1}) \text{ and } \bar{I} := \sqrt{\frac{m}{m-1}}(s + \phi^{-1} + 6\beta^{-1}).$$

By the definition of $\varrho_{n,m-1}$, this implies that

$$\mathrm{E}[h(Z^{(m)}, 6)] \le 2\varrho_{n,m-1} + \mathrm{P}(S^Y_{n,m-1} \le \bar{I}) - \mathrm{P}(S^Y_{n,m-1} \le \underline{I})$$
$$\lesssim \varrho_{n,m-1} + (\log p)^{1/2}/\phi$$

where the second line follows from the anti-concentration inequality (Lemma A.1) since $\beta^{-1} \le \phi^{-1}$. Combining presented bounds gives the claim of this step.

Step 2. We now complete the proof using the argument from [14]. Denoting

$$V_{n,m} := \sqrt{v}S^X_{n,m} + \sqrt{1-v}S^Y_{n,m},$$

we obtain

$$\begin{aligned}
\mathrm{P}(V_{n,m} \le s - \phi^{-1}) &\le \mathrm{P}(F_\beta(V_{n,m}) \le 0) \le \mathrm{E}[m(V_{n,m})] \\
&\le \mathrm{P}(F_\beta(S^W_{n,m}) \le \phi^{-1}) + (\mathrm{E}[m(V_{n,m})] - \mathrm{E}[m(S^W_{n,m})]) \\
&\le \mathrm{P}(S^W_{n,m} \le s + \phi^{-1}) + \mathrm{E}[\mathcal{I}_n] \\
&\le \mathrm{P}(S^W_{n,m} \le s - \phi^{-1}) + C(\log p)^{1/2}/\phi + \mathrm{E}[\mathcal{I}_n]
\end{aligned}$$

where the first three lines follow from the properties of $F_\beta(w)$ and $g(t)$ (recall that $m(w) = g(F_\beta(w)))$ and the fourth line from the anti-concentration inequality (Lemma A.1). Here the constant C depends only on b. Combining this chain of inequalities with the bound on $\mathrm{E}[\mathcal{I}_n]$ derived in Step 1 and noting that the bound is independent of s and that a similar argument also gives

$$\mathrm{P}(V_{n,m} \le s - \phi^{-1}) \ge \mathrm{P}(S^W_{n,m} \le s - \phi^{-1}) - C(\log p)^{1/2}/\phi - \mathrm{E}[\mathcal{I}_n]$$

completes the proof of the lemma. ∎

Proof of Corollary 5.1. Fix some rectangle $A = \{w \in \mathbb{R}^p : w_j \in [a_j, b_j]$ for all $j = 1, \dots, p\}$. For $i = 1, \dots, n$, consider random vectors $\widetilde{X}_i$ and $\widetilde{Y}_i$ in $\mathbb{R}^{2p}$ defined by $\widetilde{X}_{ij} = X_{ij}$ and $\widetilde{Y}_{ij} = Y_{ij}$ for $j = 1, \dots, p$ and $\widetilde{X}_{ij} = -X_{i,j-p}$ and $\widetilde{Y}_{ij} = -Y_{i,j-p}$ for $j = p+1, \dots, 2p$. Also, consider the constant vector s in $\mathbb{R}^{2p}$ defined by $s_j = b_j$ for $j = 1, \dots, p$ and $s_j = -a_{j-p}$ for $j = p+1, \dots, 2p$. Then

$$\mathrm{P}(S^X_{n,m} \in A) = \mathrm{P}(S^{\widetilde{X}}_{n,m} \leq s) \text{ and } \mathrm{P}(S^Y_{n,m} \in A) = \mathrm{P}(S^{\widetilde{Y}}_{n,m} \leq s)$$

where $S^{\widetilde{X}}_{n,m}$ and $S^{\widetilde{Y}}_{n,m}$ are defined as $S^X_{n,m}$ and $S^Y_{n,m}$ with X_i's and Y_i's replaced by $\widetilde{X}_i$'s and $\widetilde{Y}_i$'s. Therefore, the corollary follows by applying Lemma 5.1 to random vectors $(\widetilde{X}_i)_{i=1}^n$ and $(\widetilde{Y}_i)_{i=1}^n$ and noting that the term $M_n(\phi)$ in the lemma is replaced by $M_n(2\phi)$ in the corollary because

$$\begin{aligned}
&\mathrm{E}\left[\max_{1\leq j\leq p}|X_{ij}|^3 1\left\{\max_{1\leq j\leq p}|X_{ij}| > \sqrt{n}/(4\phi\log(2p))\right\}\right] \\
&\quad\leq \mathrm{E}\left[\max_{1\leq j\leq p}|X_{ij}|^3 1\left\{\max_{1\leq j\leq p}|X_{ij}| > \sqrt{n}/(8\phi\log(p))\right\}\right]
\end{aligned}$$

since we assume that $p \geq 2$. ∎

Appendix C. Proofs for Section 2

Proof of Theorem 2.1. The proof of this result relies on the induction Lemma 5.1 and its Corollary 5.1 stated in Section 5 and proven in Appendix B.

Let K' denote a constant from the conclusion of Corollary 5.1. This constant depends only on b. Set $K_2 := 1/(K' \vee 1)$ in (10), so that

$$\phi_n = \frac{1}{K' \vee 1}\left(\frac{\overline{L}_n^2(\log p)^4}{n}\right)^{-1/6}$$

Without loss of generality, we will assume that $\phi_n \geq 2$; otherwise, the claim of the theorem holds trivially by setting $K_1 = 2(K' \vee 1)$.

Now, applying Corollary 5.1 with $\phi = \phi_n/2$ gives for all $m = [n - \log n], \dots, n$,

$$\varrho'_{n,m} \leq \frac{\varrho'_{n,m-1}}{8(K'\vee 1)^2} + \frac{3(K'\vee 1)^2\overline{L}_n^{1/3}(\log p)^{7/6}}{n^{1/6}} + \frac{M_n(\phi_n)}{8(K'\vee 1)^2\overline{L}_n}.$$

Iterating this inequality and using inequalities $K' \vee 1 \geq 1$ and $[\log n] \leq n - [n - \log n]$ yields

$$\begin{aligned}
\varrho'_{n,n} \leq{}& \left(\frac{1}{8}\right)^{[\log n+1]} \varrho'_{n,[n-\log n]-1} \\
&+ \sum_{m=[n-\log n]}^{n}\left(\frac{1}{8}\right)^{n-m}\left(\frac{3(K'\vee 1)^2\overline{L}_n^{1/3}(\log p)^{7/6}}{n^{1/6}} + \frac{M_n(\phi_n)}{8\overline{L}_n}\right).
\end{aligned}$$

Since $\varrho'_{n,[n-\log n]-1} \leq 1$ and $\sum_{j=0}^{\infty}(1/8)^j = 8/7$, this inequality gives

$$\varrho'_{n,n} \leq \left(\frac{1}{8}\right)^{\log n} + \frac{8}{7}\left(\frac{3(K' \vee 1)^2 \overline{L}_n^{1/3}(\log p)^{7/6}}{n^{1/6}} + \frac{M_n(\phi_n)}{8\overline{L}_n}\right).$$

In addition, it follows from the assumption $n^{-1}\sum_{i=1}^n \mathrm{E}[X_{ij}^2] \geq b$ that $\overline{L}_n \geq L_n \geq c$ for some constant c that depends only on b, and so $(1/8)^{\log n} \leq (1/e)^{\log n} = 1/n \leq ((\log p)^7/n)^{1/6} \leq c^{-1/3}(\overline{L}_n^2(\log p)^7/n)^{1/6}$. Therefore, the asserted claim follows by noting that $\rho_n(\mathcal{A}^r) \leq \varrho'_{n,n}$. ∎

Proof of Corollary 2.1. The proof consists of applying Theorem 2.1. Without loss of generality, we will assume in the proof that

$$\frac{B_n^2(\log(p \vee n))^7}{n} \leq c := \min\left(b/(6A), (c'/2)^3, (K_2/2)^6\right) \tag{40}$$

where K_2 appears in (10), $A > 0$ is an absolute constant and $c' > 0$ is a constant that depends only on b; both A and c' are defined later in the proof. Otherwise, the result for both cases is trivial.

Step 1. In this step we verify that

$$\frac{1}{m}\sum_{i=1}^m \mathrm{E}[X_{ij}^2] \geq b/2 \tag{41}$$

holds for all $m = [n - \log n - 1], \ldots, n$ and $j = 1, \ldots, p$. Indeed, under the conditions (E.1) or (E.2), we have that $\mathrm{E}[X_{ij}^2] \leq AB_n^2$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$ where A is an absolute constant. Therefore,

$$\frac{1}{n}\sum_{i=m+1}^n \mathrm{E}[X_{ij}^2] \leq AB_n^2(2 + \log n)/n \leq 3AB_n^2 \log n/n \leq 3Ac$$

since $n \geq 4$. Since $c > 0$ is such that $3Ac \leq b/2$, we have

$$\frac{1}{m}\sum_{i=1}^m \mathrm{E}[X_{ij}^2] \geq \frac{1}{n}\sum_{i=1}^m \mathrm{E}[X_{ij}^2] \geq \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_{ij}^2] - b/2 \geq b/2.$$

Therefore, (41) follows.

Step 2. Given Step 1 we can apply Theorem 2.1 to get the bound on $\rho_n(\mathcal{A}^r)$ stated in Theorem 2.1:

$$\rho_n(\mathcal{A}^r) \leq K_1\left[\left(\frac{\overline{L}_n^2(\log p)^7}{n}\right)^{1/6} + \frac{M_{n,X}(\phi_n) + M_{n,Y}(\phi_n)}{\overline{L}_n}\right], \tag{42}$$

where

$$\phi_n = K_2\left(\frac{\overline{L}_n^2(\log p)^4}{n}\right)^{-1/6}$$

and where $\overline{L}_n$ is any number such that $\overline{L}_n \geq L_n$. Recall that

$$L_n = \max_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^n \mathrm{E}[|X_{ij}|^3],$$

$$M_{n,X}(\phi_n) = \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\max_{1 \leq j \leq p} |X_{ij}|^3 1\left\{ \max_{1 \leq j \leq p} |X_{ij}| > \sqrt{n}/(4\phi \log p) \right\} \right],$$

and $M_{n,Y}(\phi_n)$ is defined similarly with X_{ij}'s replaced by Y_{ij}'s.

In what follows, we need to select $\bar{L}_n$ such that $\overline{L}_n \geq L_n$ and compute upper bounds on $M_{n,X}(\phi_n)$ and $M_{n,Y}(\phi_n)$ in order to compute upper bounds on the right side of (42). The steps below carry out these computations for cases (E.1) and (E.2) separately. Inserting the resulting bounds into (42) gives the two claims of the corollary.

Case (E.1). In this case, the notation $a_n \lesssim d_n$ means that $a_n \leq C d_n$ for some constant $C > 0$ depending only on b. By condition (M.2), we have $L_n \leq B_n =: \overline{L}_n$. Observe that (E.1) implies that $\|X_{ij}\|_{\psi_1} \leq B_n$ for all i and j. Therefore, Lemma 2.2.2 in [38] shows that

$$\left\| \max_{1 \leq j \leq p} X_{ij} \right\|_{\psi_1} \leq A' B_n \log p \tag{43}$$

for some absolute constant A'. Hence, by Markov's inequality,

$$\mathrm{P}\left(\max_{1 \leq j \leq p} |X_{ij}| > x \right) \leq 2 \exp\left(-\frac{x}{A' B_n \log p} \right)$$

for all $x \geq 0$, so that applying Lemma A.8 gives

$$\begin{aligned} M_{n,X}(\phi_n) &\lesssim \left(\sqrt{n}/(\phi_n \log p) + B_n \log p\right)^3 \exp\left(-\frac{\sqrt{n}/(4\phi_n \log p)}{A' B_n \log p} \right) \\ &\lesssim \left(\sqrt{n} + B_n \log p\right)^3 \exp\left(-\frac{c' n^{1/3}}{B_n^{2/3} (\log p)^{4/3}} \right) \\ &\lesssim n^{3/2} \exp\left(-c' \log(pn) \left(\frac{B_n^2 (\log(pn))^7}{n} \right)^{-1/3} \right) \\ &\lesssim n^{3/2} \exp(-2 \log(pn)) \leq n^{-1/2} \end{aligned}$$

for some constant $c' > 0$ that depends only on b where the second, third, and fourth lines follow from (40) and the observation that

$$\frac{1}{\phi_n} = \frac{1}{K_2} \left(\frac{\overline{L}_n^2 (\log p)^4}{n} \right)^{1/6} \leq \frac{1}{K_2} \left(\frac{B_n^2 (\log(pn))^4}{n} \right)^{1/6} \leq 1 \tag{44}$$

where the last inequality again follows from (40). Further, $\mathrm{E}[Y_{ij}^2] = \mathrm{E}[X_{ij}^2] \leq A B_n^2$, so that $\|Y_{ij}\|_{\psi_1} \lesssim B_n$ by Gaussianity of Y_{ij} for all i and j. Therefore,

$M_{n,Y}(\phi_n) \lesssim n^{-1/2}$ by the same argument as that used above, and so

$$\frac{M_{n,X}(\phi_n) + M_{n,Y}(\phi_n)}{\overline{L}_n} \le M_{n,X}(\phi_n) + M_{n,Y}(\phi_n) \lesssim \left(\frac{B_n^2(\log(pn))^7}{n}\right)^{1/6}$$

where we used $B_n \ge 1$. Inserting this bound and $\bar{L}_n$ into (42) gives the required claim (12).

Case (E.2). In this case, the notation $a_n \lesssim d_n$ means that $a_n \le Cd_n$ for some constant $C > 0$ depending only on b and q. Without loss of generality, in addition to (40), we will assume that

$$\frac{B_n^2(\log p)^{3/2}}{n^{1/2-1/q}} \le (K_2/2)^{3/2}; \tag{45}$$

otherwise, the result for the (E.2) case is trivial. Further, by definition of L_n and condition on B_n given in (M.2) we have that:

$$L_n \le B_n \le \left(B_n + \frac{B_n^2}{(\log p)^{1/2} n^{1/2-2/q}}\right) =: \overline{L}_n.$$

Therefore, since $(a+d)^{1/6} \le a^{1/6} + d^{1/6}$ for any $a, d \ge 0$,

$$\left(\frac{\overline{L}_n^2(\log p)^7}{n}\right)^{1/6} \le \left(\frac{B_n^2(\log p)^7}{n}\right)^{1/6} + \left(\frac{B_n(\log p)^{3/2}}{n^{1/2-1/q}}\right)^{2/3}.$$

Hence, under (40) and (45), we have that that

$$\frac{1}{\phi_n} = \frac{1}{K_2}\left(\frac{\overline{L}_n^2(\log p)^4}{n}\right)^{1/6} \le \frac{1}{K_2}\left(\frac{\overline{L}_n^2(\log p)^7}{n}\right)^{1/6} \le 1.$$

Next, using the bound $\mathrm{E}[|Z|^3 1(|Z| > t)] \le \mathrm{E}[|Z|^3(|Z|/t)^{q-3} 1(|Z| > t)] \le \mathrm{E}[|Z|^q t^{3-q}]$ holding for any random variable Z and $t \ge 0$, we conclude that

$$M_{n,X}(\phi_n) \lesssim \frac{B_n^q \phi^{q-3}(\log p)^{q-3}}{n^{q/2-3/2}}.$$

Using an elementary inequality, we also conclude that

$$\frac{1}{\overline{L}_n} \lesssim \frac{(\log p)^{1/2} n^{1/2-2/q}}{B_n^2} \tag{46}$$

Hence, using the bound

$$\phi_n = K_2\left(\frac{\overline{L}_n^2(\log p)^4}{n}\right)^{-1/6} \lesssim \frac{n^{1/3-2/(3q)}}{B_n^{2/3}(\log p)^{1/2}}$$

which follows from (46), gives us

$$M_{n,X}(\phi_n) \lesssim \frac{B_n^{q/3+2}(\log p)^{q/2-3/2}}{n^{q/6+1/6-2/q}} \text{ and } \frac{M_{n,X}(\phi_n)}{\overline{L}_n} \lesssim \frac{1}{\log p}\left(\frac{B_n(\log p)^{3/2}}{n^{1/2-1/q}}\right)^{q/3}.$$

Moreover, as in the proof for the (E.1) case, we have that

$$\frac{M_{n,Y}(\phi_n)}{\bar{L}_n} \lesssim \left(\frac{B_n^2(\log(p \vee n))^7}{n}\right)^{1/6}.$$

Inserting these inequalities and $\bar{L}_n$ into (42) gives the required claim (13). ∎

Appendix D. Proofs for Section 3

Proof of Corollary 3.1. Consider the (E.1′) case first. Fix any $A \in \mathcal{A}^s$. Let A^m be an approximating m-generated convex set as in (C.1). By assumption, $A^m \subset A \subset A^{m,\epsilon}$, and so denoting

$$\rho := |\mathrm{P}(S_n^X \in A^m) - \mathrm{P}(S_n^Y \in A^m)| \vee |\mathrm{P}(S_n^X \in A^{m,\epsilon}) - \mathrm{P}(S_n^Y \in A^{m,\epsilon})|,$$

we obtain

$$\begin{aligned}\mathrm{P}(S_n^X \in A) &\le \mathrm{P}(S_n^X \in A^{m,\epsilon}) \le \mathrm{P}(S_n^Y \in A^{m,\epsilon}) + \rho \\ &\le \mathrm{P}(S_n^Y \in A^m) + C\epsilon(\log p)^{1/2} + \rho \\ &\le \mathrm{P}(S_n^Y \in A) + C\epsilon(\log p)^{1/2} + \rho.\end{aligned}$$

An analogous argument also gives

$$\mathrm{P}(S_n^X \in A) \ge \mathrm{P}(S_n^Y \in A) - C\epsilon(\log p)^{1/2} - \rho.$$

Therefore,

$$\left|\mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A)\right| \le C\epsilon(\log p)^{1/2} + \rho.$$

Further,

$$\epsilon(\log p)^{1/2} \le C\frac{(\log p)^{1/2}}{n} \le C\left(\frac{B_n^2(\log(pn))^7}{n}\right)^{1/6}$$

since $B_n \ge 1$ and $\epsilon \le a/n$. Therefore, the asserted claim in the (E.1′) case follows by noting that

$$\rho \le C\left(\frac{B_n^2(\log(pn))^7}{n}\right)^{1/6}$$

under (E.1′) by Corollary 2.1 applied to vectors $\widetilde{X}_1, \dots, \widetilde{X}_n$ and $\widetilde{Y}_1, \dots, \widetilde{Y}_n$ where the latter sequence is defined by analogy with the former sequence but using vectors $Y_1, \dots, Y_n$. The asserted claim in the (E.2′) case follows from the same argument. This completes the proof of the corollary. ∎

Proof of Corollary 3.2. Since X_i is a centered random vector with a log-concave distribution in $\mathbb{R}^p$, Borell's inequality [9, Lemma 3.1] implies that $\|v'X_i\|_{\psi_1} \leq c(\mathrm{E}[(v'X_i)^2])^{1/2}$ for all $v \in \mathbb{R}^p$ for some universal constant $c > 0$ [see 25, Appendix III]; hence if the maximal eigenvalue of each $\mathrm{E}[X_iX_i']$ is bounded by a constant K, then any deformed rectangle $A \in \mathcal{A}^s$ obeys conditions (M.2′) and (E.1′) with B_n replaced by a constant C depending only on c and K. Therefore, the asserted claim follows by applying Corollary 3.1. ■

Proof of Lemma 3.1. For convex sets P_1 and P_2 containing the origin and such that $P_1 \subset P_2$, denote

$$d_{BM}(P_1, P_2) := \inf\{\epsilon : P_2 \subset (1+\epsilon)P_1\}.$$

It is immediate to verify that the function d_{BM} has the following useful property: for any convex sets P_1, P_2, P_3, and P_4 containing the origin and such that $P_1 \subset P_2$ and $P_3 \subset P_4$,

$$d_{BM}(P_1 \cap P_3, P_2 \cap P_4) \leq d_{BM}(P_1 \cap P_2) \vee d_{BM}(P_3 \cap P_4). \tag{47}$$

Let $A = \cap_{q=1}^{Q} A_q$ be a sparsely representation of A as appeared in the statement of the lemma. Fix any A_q. By assumption, the indicator function $w \mapsto I(w \in A_q)$ depends only on $s_q \leq s$ components of its argument $w = (w_1, \dots, w_p)$. Let E_q denote the subspace of $\mathbb{R}^p$ corresponding to these s_q components. Since A contains the origin, it follows that A_q contains the origin as well. Therefore, applying Corollary 1.5 in [3] shows that one can construct a polytope $P_q' \subset E_q$ with at most $(\gamma((\mu+1)/\epsilon)^{1/2}\log(1/\epsilon))^{s_q}$ vertices such that the section $A_q \cap E_q$ satisfies

$$P_q' \subset A_q \cap E_q \subset (1+\epsilon)P_q'.$$

Clearly, the polytope P_q' has

$$m_q \leq \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s_q^2} \leq \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s^2} \tag{48}$$

faces (of dimension $s_q - 1$). Indeed, since E_q is a s_q-dimensional vector space, a polytope with k vertices has no more than k^{s_q} faces. Hence, one can construct an m_q-generated convex set P_q such that $P_q \subset A_q \subset (1+\epsilon)P_q$ and all vectors in $\mathcal{V}(P_q)$ having at most s non-zero components. Hence,

$$d_{BM}(P_q, A_q) \leq \epsilon.$$

Next, it follows from (47) that

$$d_{BM}(\cap_{q=1}^{Q} P_q, \cap_{q=1}^{Q} A_q) \leq \epsilon.$$

Therefore, defining $A^m = \cap_{q=1}^{Q} P_q$, we obtain from $A = \cap_{q=1}^{Q} A_q$ that

$$A^m \subset A \subset (1+\epsilon)A^m \subset A^{m,R\epsilon}$$

where the last assertion follows from the assumption that $\sup_{w\in A}|w|\leq R$. Since A^m is an m-generated convex set with $m\leq\sum_{q=1}^{Q}m_q$, the asserted claim of the lemma now follows from (48). ■

Proof of Lemma 3.2. Fix any $v=(v_1,\dots,v_p)'\in\mathcal{V}(A^m)$. To establish (M.2′), observe that by construction of A^m in Lemma 3.1, v has at most s non-zero components. Let $J(v)$ be the set containing positions of non-zero components, so that $|J(v)|\leq s$. Using an elementary inequality $(\sum_{j\in J(v)}|a_j|)^{2+k}\leq s^{2+k}\sum_{j\in J(v)}|a_j|^{2+k}$ for all $a=(a_1,\dots,a_p)'\in\mathbb{R}^p$, we obtain for $k=1$ or 2,

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[|v'X_i|^{2+k}]\leq\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big[\Big(\sum_{j\in J(v)}|X_{ij}|\Big)^{2+k}\Big]$$
$$\leq s^{2+k}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big[\sum_{j\in J(v)}|X_{ij}|^{2+k}\Big]\leq s^{3+k}B_n^k\leq(B_n')^k$$

where $B_n'=s^4B_n$. This gives condition (M.2′).

Next, under (E.1′), we have $\|X_{ij}\|_{\psi_1}\leq B_n$. Therefore, by the triangle inequality, $\|v'X_i\|_{\psi_1}\leq\sum_{j\in J(v)}\|X_{ij}\|_{\psi_1}\leq sB_n$ showing that the vectors $\widetilde{X}_i$, $i=1,\dots,n$, satisfy (E.1′) with B_n replaced by sB_n.

Finally, under (E.2′),

$$\mathrm{E}\left[\max_{v\in\mathcal{V}(A^m)}|v'X_i|^q\right]\leq s^q\mathrm{E}\left[\max_{1\leq j\leq p}|X_{ij}|^q\right],$$

showing that the vectors $\widetilde{X}_i$, $i=1,\dots,n$, satisfy (E.2′) with B_n replaced by sB_n. This completes the proof of the lemma. ■

Appendix E. Proofs for Section 4

Proof of Theorem 4.1. We first show that

$$\varrho_n^{MB}:=\sup_{s\in\mathbb{R}^p}\left|\mathrm{P}(S_n^{eX}\leq s\mid X_1^n)-\mathrm{P}(S_n^Y\leq s)\right|\lesssim\Delta_n^{1/3}(\log p)^{2/3}\qquad(49)$$

up to a constant C that depends only on b. To show (49), fix $s=(s_1,\dots,s_p)'\in\mathbb{R}^p$. As in the proof of Lemma 5.1, for $\beta>0$ and $w\in\mathbb{R}^p$, define

$$F_\beta(w):=\frac{1}{\beta}\log\left(\sum_{j=1}^{p}\exp(\beta(w_j-s_j))\right).$$

Note that conditional on X_1^n, S_n^{eX} is a zero-mean Gaussian random vector with covariance matrix Σ_n^{eX}. Hence, using the same argument as that in the proof of Theorem 1 in [17], we obtain for any $g\in C^2(\mathbb{R})$ with $\|g'\|_\infty\vee\|g''\|_\infty<\infty$ that

$$|\mathrm{E}[g(F_\beta(S_n^{eX}))\mid X_1^n]-\mathrm{E}[g(F_\beta(S_n^Y))]|\leq(\|g''\|_\infty/2+\beta\|g'\|_\infty)\Delta_n$$

(Formally, the proof of Theorem 1 in [17] imposes $s = 0$ but it is easy to verify that their proof also applies for all $s \in \mathbb{R}^p$.) Therefore, as in Step 2 of the proof of Lemma 5.1, we obtain with $\phi = \beta/\log p$ that

$$|\mathrm{P}(S_n^{eX} \le s - \phi \mid X_1^n) - \mathrm{P}(S_n^Y \le s - \phi)| \lesssim \frac{(\log p)^{1/2}}{\phi} + (\phi^2 + \beta\phi)\Delta_n$$

up to a constant C that depends only on b. Substituting $\beta = \phi \log p$, optimizing the resulting expression with respect to ϕ, and noting that s is arbitrary give (49).

The conclusion of the theorem now follows by noting that on the event $\Delta_n \le \bar{\Delta}_n$, we have $\varrho_n^{MB} \lesssim \bar{\Delta}_n^{1/3}(\log p)^{2/3}$, and applying the same argument as that in the proof of Corollary 5.1. ∎

Proof of Corollary 4.1. In this proof, c and C are constants that depend only on b under (E.1′) and on b and q under (E.2′) but their values may change at each appearance. Also, for brevity of notation, in this proof we implicitly assume that i is varying over $\{1, \dots, n\}$ and j and k are varying over $\{1, \dots, p\}$. Finally, without loss of generality, we will assume that

$$\frac{B_n^2(\log(pn))^5(\log(1/\alpha))^2}{n} \le 1; \tag{50}$$

otherwise, the asserted claims are trivial.

Fix any $A \in \mathcal{A}^s$. Let A^m be an approximating m-generated convex set as in (C). By assumption, $A^m \subset A \subset A^{m,\epsilon}$. Denote

$$\rho := \max\Big\{|\mathrm{P}(S_n^{eX} \in A^m \mid X_1^n) - \mathrm{P}(S_n^Y \in A^m)|,\\ |\mathrm{P}(S_n^{eX} \in A^{m,\epsilon} \mid X_1^n) - \mathrm{P}(S_n^Y \in A^{m,\epsilon})|\Big\}.$$

As in the proof of Corollary 3.1,

$$\begin{aligned}|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| &\le C\epsilon(\log p)^{1/2} + \rho\\ &\le C\left(\frac{B_n^2(\log(pn))^7}{n}\right)^{1/6} + \rho.\end{aligned}$$

Therefore, the problem reduces to the case of rectangles $\mathcal{A} = \mathcal{A}^r$; that is, it suffices to prove the bounds (19) and (20) with $\rho_n^{MB}(\mathcal{A}^s)$ replaced by $\rho_n^{MB}(\mathcal{A}^r)$ and conditions (C), (M.1′), (M.2′), and (E.1′) (or (E.2′)) replaced by (M.1), (M.2), and (E.1) (or (E.2), respectively). For the latter problem, we will apply Theorem 4.1.

Note that $\mathrm{E}[X_iX_i'] = \mathrm{E}[Y_iY_i']$ for all i. Therefore,

$$\Sigma_n^{eX} - \Sigma_n^Y = \frac{1}{n}\sum_{i=1}^n (X_iX_i' - \mathrm{E}[X_iX_i']) - \widehat{\mu}_n^X(\widehat{\mu}_n^X)'.$$

Hence, by the triangle inequality,

$$\Delta_n \le \Delta_{n,1} + \Delta_{n,2}^2 \tag{51}$$

where

$$\Delta_{n,1} := \max_{1 \le j,k \le p} \left| \frac{1}{n} \sum_{i=1}^{n} (X_{ij}X_{ik} - \mathrm{E}[X_{ij}X_{ik}]) \right|, \tag{52}$$

$$\Delta_{n,2} := \left(\max_{1 \le j,k \le p} \left| \widehat{\mu}_{nj}^{X} \widehat{\mu}_{nk}^{X} \right| \right)^{1/2} = \max_{1 \le j \le p} \left| \widehat{\mu}_{nj}^{X} \right|. \tag{53}$$

The asserted claims follow from the bounds on $\Delta_{n,1}$ and $\Delta_{n,2}$, derived separately for (E.1) and (E.2) cases below, and Theorem 4.1.

Case (E.1). We start with some preliminary calculations. We have

$$\sigma_n^2 := \max_{j,k} \sum_{i=1}^{n} \mathrm{E}\left[(X_{ij}X_{ik} - \mathrm{E}[X_{ij}X_{ik}])^2 \right] \le \max_{j,k} \sum_{i=1}^{n} \mathrm{E}[|X_{ij}X_{ik}|^2] \tag{54}$$

$$\le \max_{j,k} \left(\sum_{i=1}^{n} \mathrm{E}[|X_{ij}|^4] \sum_{i=1}^{n} \mathrm{E}[|X_{ik}|^4] \right)^{1/2} \le nB_n^2 \tag{55}$$

where the second line follows from Hölder's inequality and (M.2). In addition,

$$\| \max_{i,j,k} |X_{ij}X_{ik}| \|_{\psi_{1/2}} = \| \max_{i,j} |X_{ij}|^2 \|_{\psi_{1/2}} = \| \max_{i,j} |X_{ij}| \|_{\psi_1}^2 \lesssim (B_n \log(pn))^2$$

by (E.1). Thus, denoting $M_n := \max_{i,j,k} |X_{ij}X_{ik} - \mathrm{E}[X_{ij}X_{ik}]|$, we obtain

$$\begin{aligned} \|M_n\|_{\psi_{1/2}} &\lesssim \| \max_{i,j,k} |X_{ij}X_{ik}| \|_{\psi_{1/2}} + \max_{i,j,k} \mathrm{E}[|X_{ij}X_{ik}|] \\ &\lesssim (B_n \log(pn))^2 + B_n^2 \lesssim (B_n \log(pn))^2, \end{aligned}$$

which also implies that $(\mathrm{E}[M_n^2])^{1/2} \lesssim (B_n \log(pn))^2$. So Lemma A.2 yields

$$\begin{aligned} \mathrm{E}[\Delta_{n,1}] &\lesssim \frac{\sqrt{\sigma_n^2 \log p}}{n} + \frac{\sqrt{\mathrm{E}[M_n^2]} \log p}{n} \\ &\lesssim \left(\frac{B_n^2 \log p}{n} \right)^{1/2} + \frac{B_n^2 (\log(pn))^3}{n} \lesssim \left(\frac{B_n^2 \log(pn)}{n} \right)^{1/2} \end{aligned}$$

where the last inequality follows from (50). Thus, applying Lemma A.3 (i) with $\beta = 1/2$ yields for any $t > 0$,

$$\begin{aligned} &\mathrm{P}\left(\Delta_{n,1} > C \left(\frac{B_n^2 \log(pn)}{n} \right)^{1/2} + t \right) \\ &\quad \lesssim \exp\left(-\frac{cnt^2}{B_n^2} \right) + \exp\left(-\frac{c\sqrt{nt}}{B_n \log(pn)} \right) \end{aligned}$$

for sufficiently large $C > 0$ and sufficiently small $c > 0$. Setting $t = C(B_n^2 \log(pn)(\log(1/\alpha))^2/n)^{1/2}$ with sufficiently large $C > 0$, this inequality yields

$$\mathrm{P}\left(\Delta_{n,1} > C\left(\frac{B_n^2 \log(pn)(\log(1/\alpha))^2}{n}\right)^{1/2}\right)$$
$$\leq \frac{\alpha}{4} + \frac{1}{4}\exp\left(-\frac{n^{1/4}(\log(1/\alpha))^{1/2}}{B_n^{1/2}(\log(pn))^{3/4}}\right) \leq \frac{\alpha}{2}$$

by (50). Hence,

$$\mathrm{P}\left((\Delta_{n,1}(\log p)^2)^{1/3} > C\left(\frac{B_n^2(\log(pn))^5(\log(1/\alpha))^2}{n}\right)^{1/6}\right) \leq \frac{\alpha}{2}$$

It is also easy to check that the same inequality holds with $\Delta_{n,1}$ replaced by $\Delta_{n,2}^2$. Therefore, the asserted claim for the (E.1) case follows from Theorem 4.1 applied with $\bar{\Delta}_n = \Delta_{n,1} + \Delta_{n,2}^2$.

Case (E.2). Define σ_n^2 and M_n by the same expressions as those in the (E.1) case. Then the bounds (54) and (55) on σ_n^2 hold under (E.2) as well. For the bound on M_n, using an elementary inequality $|x-y|^{q/2} \lesssim |x|^{q/2} + |y|^{q/2}$ for all $x, y \in \mathbb{R}$ and $q \geq 2$, we obtain

$$\mathrm{E}[M_n^{q/2}] \lesssim \mathrm{E}[\max_{i,j,k} |X_{ij}X_{ik}|^{q/2}] + \max_{i,j,k} (\mathrm{E}[|X_{ij}X_{ik}|])^{q/2}$$
$$\lesssim \mathrm{E}[\max_{i,j,k} |X_{ij}X_{ik}|^{q/2}] = \mathrm{E}[\max_{i,j} |X_{ij}|^q] \lesssim nB_n^q$$

where the third line follows from Jensen's inequality and (E.2). The last bound also implies that $(\mathrm{E}[M_n^2])^{1/2} \lesssim n^{2/q}B_n^2$. So Lemma A.2 yields

$$\mathrm{E}[\Delta_{n,1}] \lesssim \frac{\sqrt{\sigma_n^2 \log p}}{n} + \frac{\sqrt{\mathrm{E}[M_n^2]}\log p}{n} \lesssim \left(\frac{B_n^2 \log p}{n}\right)^{1/2} + \frac{B_n^2 \log p}{n^{1-2/q}}.$$

Thus, applying Lemma A.3 (ii) with $s = q/2$ yields for any $t > 0$,

$$\mathrm{P}\left(\Delta_{n,1} > C\left(\frac{B_n^2 \log p}{n}\right)^{1/2} + \frac{CB_n^2 \log p}{n^{1-2/q}} + t\right)$$
$$\lesssim \exp\left(-\frac{cnt^2}{B_n^2}\right) + \frac{B_n^q}{t^{q/2}n^{q/2-1}}$$

for sufficiently large $C > 0$ and sufficiently small $c > 0$. Setting

$$t = C\left[\left(\frac{B_n^2 \log(pn)(\log(1/\alpha))^2}{n}\right)^{1/2} + \frac{B_n^2}{\alpha^{2/q}n^{1-2/q}}\right]$$

with sufficiently large $C > 0$ yields

$$\mathrm{P}\left(\Delta_{n,1} > C\left(\frac{B_n^2 \log(pn)(\log(1/\alpha))^2}{n}\right)^{1/2} + \frac{CB_n^2 \log p}{\alpha^{2/q}n^{1-2/q}}\right) \leq \frac{\alpha}{4} + \frac{\alpha}{4} = \frac{\alpha}{2}$$

by (50). Hence, using an elementary inequality $|x+y|^{1/3} \leq |x|^{1/3}+|y|^{1/3}$ for all $x, y \in \mathbb{R}$, we obtain that the probability that $(\Delta_{n,1}(\log p)^2)^{1/3}$ is bounded from below by

$$C\left[\left(\frac{B_n^2(\log(pn))^5(\log(1/\alpha))^2}{n}\right)^{1/6} + \left(\frac{B_n^2(\log p)^3}{\alpha^{2/q}n^{1-2/q}}\right)^{1/3}\right]$$

is bounded from above by $\alpha/2$. It is also easy to check that the same inequality holds with $\Delta_{n,1}$ replaced by $\Delta_{n,2}^2$. Therefore, the asserted claim for the (E.2) case follows from Theorem 4.1. This completes the proof of the corollary. ■

Proof of Theorem 4.2. By the triangle inequality,

$$\rho_n^{EB} \leq \rho_n^{MB} + \varrho_n^{EB}$$

where

$$\varrho_n^{EB} := \sup_{A\in\mathcal{A}^r}\left|\mathrm{P}(S_n^{X^*} \in A \mid X_1^n) - \mathrm{P}(S_n^{eX} \in A \mid X_1^n)\right|.$$

Also, conditional on X_1^n, $X_1^* - \widehat{\mu}_n^X, \dots, X_n^* - \widehat{\mu}_n^X$ are i.i.d. with zero mean and covariance Σ_n^{eX}. In addition, conditional on X_1^n,

$$S_n^{eX} =_d \frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i^*$$

where $Y_1^*, \dots, Y_n^*$ are i.i.d. zero-mean Gaussian random vectors with the same covariance Σ_n^{eX}. Therefore, the result follows by applying Theorem 2.1 conditional on X_1^n (with L_n and $M_n(\phi_n)$ in Theorem 2.1 substituted by $\widehat{L}_n$ and $\widehat{M}_n(\phi_n)$) to bound ϱ_n^{EB} on the event that $\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj})^2] \geq b$ for all $1 \leq j \leq p$, $\overline{L}_n \geq \widehat{L}_n$, and $\overline{M}_n \geq \widehat{M}_n(\phi_n)$. ■

Proof of Corollary 4.2. Here c, C are constants depending only on b, q, K; their values may change from place to place. We first note that, for sufficiently small $c > 0$, we may assume that

$$B_n^2(\log(pn))^5(\log(1/\alpha))^2 \leq cn, \tag{56}$$

since otherwise we can make the assertion of the lemma trivial by setting C sufficiently large.

Further, by the same argument as that used in the proof of Corollary 4.1, the problem reduces to the case of rectangles $\mathcal{A} = \mathcal{A}^r$; that is, it suffices to prove the bounds (22) and (23) with $\rho_n^{EB}(\mathcal{A}^s)$ replaced by $\rho_n^{EB}(\mathcal{A}^r)$ and conditions (C), (M.1′), (M.2′), and (E.1′) (or (E.2′)) replaced by (M.1), (M.2), and (E.1) (or (E.2), respectively). For the latter problem, we will apply Theorem 4.2.

Case (E.1). With (56) in mind, by the proof of Corollary 4.1, we see that $\mathrm{P}(\Delta_n > b/2) \leq \alpha/6$, so that with probability larger than $1 - \alpha/6$, $b/2 \leq \mathbb{E}_n[(X_{ij} - \widehat{\mu}_{ij}^X)^2] \leq CB_n$ for all $j = 1, \dots, p$. We turn to bounding

$\widehat{L}_n$. Using the inequality $|a-b|^3 \le 4(|a|^3+|b|^3)$ together with Jensen's inequality, we have

$$\widehat{L}_n \le 4(\max_{1\le j\le p} \mathbb{E}_n[|X_{ij}|^3] + \max_{1\le j\le p} |\widehat{\mu}^X_{nj}|^3) \le 8 \max_{1\le j\le p} \mathbb{E}_n[|X_{ij}|^3].$$

By Lemma A.4,

$$\begin{aligned}\mathrm{E}[\max_{1\le j\le p} \mathbb{E}_n[|X_{ij}|^3]] &\le C(L_n + n^{-1}\mathrm{E}[\max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|^3]\log p) \\ &\le C(B_n + n^{-1}B_n^3(\log(pn))^4).\end{aligned}$$

Note that $\| |X_{ij}|^3 \|_{\psi_{1/3}} \le \|X_{ij}\|_{\psi_1}^3 \le B_n^3$, so that applying Lemma A.5 (i) with $\beta = 1/3$, we have for every $t>0$,

$$\mathrm{P}\{\widehat{L}_n \ge C(B_n + n^{-1}B_n^3(\log(pn))^4 + n^{-1}B_n^3 t^3)\} \le 3e^{-t}.$$

Taking $t = \log(18/\alpha) \le C\log(pn)$, we conclude that, with $\overline{L}_n = CB_n$ (recall (56)),

$$\mathrm{P}(\widehat{L}_n > \overline{L}_n) \le \alpha/6.$$

Consider to bound $\widehat{M}_{n,X}(\phi_n)$. Observe that

$$\max_{1\le j\le p} |X_{ij} - \widehat{\mu}^X_{nj}| \le 2\max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|,$$

so that

$$\mathrm{P}\{\widehat{M}_{n,X}(\phi_n) > 0\} \le \mathrm{P}\{\max_{i,j}|X_{ij}| > \sqrt{n}/(8\phi_n\log p)\}.$$

Since $\|X_{ij}\|_{\psi_1} \le B_n$, the right side is bounded by

$$2(pn)\exp\{-\sqrt{n}/(8B_n\phi_n\log p)\}.$$

Observe that

$$B_n\phi_n\log p \le Cn^{-1/6}B_n^{2/3}(\log p)^{1/3},$$

so that using (56), we conclude that $\mathrm{P}\{\widehat{M}_{n,X}(\phi_n) > 0\} \le \alpha/6$. For $\widehat{M}_{n,Y}(\phi_n)$, since with probability larger than $1-\alpha/6$, $\mathbb{E}_n[(X_{ij} - \widehat{\mu}^X_{nj})^2] \le CB_n$ for all $j = 1,\dots,p$, on that event, conditional on $X_1,\dots,X_n$, $\|S^{eX}_{nj}\|_{\psi_2} \le CB_n^{1/2}$ for all $j = 1,\dots,p$. Hence, using the same argument used in bounding $\widehat{M}_{n,X}(\phi_n)$, we conclude that

$$\mathrm{P}\{\widehat{M}_{n,Y}(\phi_n) > 0\} \le \alpha/6 + \alpha/6 = \alpha/3,$$

which implies that

$$\mathrm{P}\{\widehat{M}_n(\phi_n) = 0\} > 1 - (\alpha/6 + \alpha/3) = 1 - \alpha/2.$$

Taking these together, by Theorem 4.2, with probability larger than $1 - (\alpha/6 + \alpha/6 + \alpha/2) = 1 - 5\alpha/6$, we have

$$\rho_n^{EB} \le \rho_n^{MB} + C\{n^{-1}B_n^2(\log(pn))^7\}^{1/6}.$$

The final conclusion follows from Corollary 4.1.

Case (E.2). In this case, in addition to (56), we may assume that

$$\frac{B_n^2(\log(pn))^3}{\alpha^{2/q}n^{1-2/q}} \leq c \leq 1, \tag{57}$$

since otherwise we can make the assertion of the lemma trivial by setting C sufficiently large. Then as in the previous case, by the proof of Corollary 4.1, with probability larger than $1-\alpha/6$, $b/2 \leq \mathbb{E}_n[(X_{ij}-\widehat{\mu}_{nj}^X)^2] \leq CB_n$ for all $j=1,\dots,p$.

To bound $\widehat{L}_n$, recall that $\widehat{L}_n \leq 8\max_{1\leq j\leq p}\mathbb{E}_n[|X_{ij}|^3]$, and by Lemma A.4,

$$\mathrm{E}[\max_{1\leq j\leq p}\mathbb{E}_n[|X_{ij}|^3]] \leq C(B_n + B_n^3 n^{-1+3/q}\log p).$$

Hence by applying Lemma A.5 (ii) with $s=q/3$, we have for every $t>0$,

$$\mathrm{P}\{\widehat{L}_n \geq C(B_n+B_n^3n^{-1+3/q}\log p)+n^{-1}t\} \leq Ct^{-q/3}\mathrm{E}[\max_{i,j}|X_{ij}|^q] \leq Ct^{-q/3}nB_n^q.$$

Solving $Ct^{-q/3}nB_n^q = \alpha/6$, we conclude that

$$\mathrm{P}(\widehat{L}_n \geq \overline{L}_n) \leq \alpha/6,$$

where $\overline{L}_n = C(B_n + B_n^3 n^{-1+3/q}\alpha^{-3/q}\log p)$.

We turn to bounding $\widehat{M}_{n,X}(\phi_n)$. As in the previous case,

$$\mathrm{P}\{\widehat{M}_{n,X}(\phi_n)>0\} \leq \mathrm{P}\{\max_{i,j}|X_{ij}| > \sqrt{n}/(8\phi_n\log p)\}.$$

Since the right side is nondecreasing in ϕ_n, and

$$\phi_n \leq cB_n^{-1}n^{1/2-1/q}\alpha^{1/q}(\log p)^{-1},$$

we have (by choosing the constant C in $\overline{L}_n$ large enough)

$$\mathrm{P}\{\max_{i,j}|X_{ij}| > \sqrt{n}/(8\phi_n\log p)\} \leq n\mathrm{P}\{\max_{j}|X_{ij}| > CB_n n^{1/q}\alpha^{-1/q}\} \leq \alpha/6.$$

For $\widehat{M}_{n,Y}(\phi_n)$, we make use of the argument in the previous case, and conclude that

$$\mathrm{P}\{\widehat{M}_{n,Y}(\phi_n)>0\} \leq \alpha/2.$$

The rest of the proof is the same as in the previous case. Note that

$$\left(\frac{\overline{L}_n^2(\log(pn))^7}{n}\right)^{1/6} \leq C\left[\left(\frac{B_n^2(\log(pn))^7}{n}\right)^{1/6} + \left(\frac{B_n^2(\log(pn))^3}{\alpha^{2/q}n^{1-2/q}}\right)^{1/2}\right],$$

and because of (57), the second term inside the bracket on the right-hand side is at most

$$\left(\frac{B_n^2(\log(pn))^3}{\alpha^{2/q}n^{1-2/q}}\right)^{1/3}.$$

This completes the proof of the corollary. ∎

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000-1034.

[2] Adamczak, R. (2010). A few remarks on the operator norm of random Toeplitz matrices. *J. Theoret. Probab.* **23** 85-108.

[3] Barvinok, A. (2013). Thrifty approximations of convex bodies by polytopes. *Int. Math. Res. Not.* **00** 1-16.

[4] Bentkus, V. (1985). Lower bounds for the rate of convergence in the central limit theorem in Banach spaces. *Lithuanian Mathematical Journal* **25** 312-320.

[5] Bentkus, V. (1986). Dependence of the Berry-Esseen estimate on the dimension (in Russian). *Litovsk. Mat. Sb.* **26** 205-210.

[6] Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *J. Stat. Plann. Infer.* **113** 385-402.

[7] Bhattacharya, R. (1975). On the errors of normal approximation. *Ann. Probab.* **3** 815-828.

[8] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379-386.

[9] Borell, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239-252.

[10] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

[11] Chatterjee, S. (2005a). A simple invariance theorem. arXiv:math/0508213.

[12] Chatterjee, S. (2006). A generalization of Lindeberg's principle. *Ann. Probab.* **34** 2061-2076.

[13] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257-283.

[14] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786-2819.

[15] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Supplemental Material for Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786-2819.

[16] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014a). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564-1597.

[17] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014b). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. To appear in *Probab. Theory Related Fields.*

[18] Dudley, R. (2002). *Uniform Central Limit Theorems.* Cambridge University Press.

[19] Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.* **360** 6677-6693.

[20] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724-739.

[21] Goldstein, L., and Rinott, Y. (1996) Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1-17.

[22] Klivans, A., O'Donnell, R., and Servedio, R. (2008). Learning geometric concepts via Gaussian surface area. *49th Annual IEEE Symposium on Foundations of Computer Science.*

[23] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer.

[24] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863-884.

[25] Milman, V.D. and Shechtman, G. (1986). *Asymptotic Theory of Finite Dimensional Normed Spaces.* Lecture Notes in Mathematics 1200 Springer.

[26] Nagaev, S. (1976). An estimate of the remainder term in the multidimensional central limit theorem. *Proc. Third Japan-USSR Symp. Probab. Theory.* **550** 419-438.

[27] Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In: *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics Volume 1807, Springer, pp. 169-187.

[28] Panchenko, D. (2013). *The Sherrington-Kirkpatrick Model.* Springer.

[29] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150-2173.

[30] Röllin, A. (2011). Stein's method in high dimensions with applications, *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 529-549.

[31] Sazonov, V. (1968). On the multi-dimensional central limit theorem. *Sanhya Ser. A* **30** 181-204.

[32] Sazonov, V. (1981). *Normal Approximations: Some Recent Advances.* Lecture Notes in Mathematics Volume 879, Springer.

[33] Senatov, V. (1980). Several estimates of the rate of convergence in the multidimensional CLT. *Dokl. Acad. Nauk SSSR* **254** 809-812.

[34] Sweeting, T. (1977). Speed of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28-41.

[35] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463-501.

[36] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.

[37] Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians.* Springer.

[38] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer.

(V. Chernozhukov) Department of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.
E-mail address: vchern@mit.edu

(D. Chetverikov) Department of Economics, UCLA, Bunche Hall, 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.
E-mail address: chetverikov@econ.ucla.edu

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo 113-0033, Japan.
E-mail address: kkato@e.u-tokyo.ac.jp