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 coordinates 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 by X_{ij} the j-th coordinate of X_i, so that $X_i = (X_{i1},\dots,X_{ip})'$. We 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

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

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, $Y_i \sim N(0, \mathrm{E}[X_i X_i'])$. Define the normalized sum for the Gaussian random vectors:

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

Date: First version: February 07, 2014. This version: March 10, 2015.

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

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

We are interested in bounding the quantity

$$\rho_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} |\mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A)|, \tag{1}$$

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, [6, 7, 8, 21, 27, 33, 34, 35, 36] (see [16] 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, for I being the $p \times p$ identity matrix, Bentkus [7] established one of the sharpest results in this direction and proved that when $X_1, \dots, X_n$ are i.i.d. with $\mathrm{E}[X_i X_i'] = I$,

$$\rho_n(\mathcal{A}) \leq c_p(\mathcal{A}) \frac{\mathrm{E}[\|X_i\|^3]}{\sqrt{n}}, \tag{2}$$

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 $c_p(\mathcal{A}) \leq 400 p^{1/4}$ 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 $\mathrm{E}[X_i X_i'] = I$, $\mathrm{E}[\|X_i\|^3] \geq (\mathrm{E}[\|X_i\|^2])^{3/2} = p^{3/2}$, and hence in order to make the right hand side of (2) 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 [27] that $\rho_n(\mathcal{A}) \geq c\mathrm{E}[\|X_i\|^3]/\sqrt{n}$ 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

$$\rho_n(\mathcal{A}) \to 0 \text{ even if } p \text{ is potentially larger or much larger than } n. \tag{3}$$

In this paper, we derive bounds on $\rho_n(\mathcal{A})$ for $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$ being the class of all rectangles, or more generally for $\mathcal{A} = \mathcal{A}^{\mathrm{si}}$ being the class of simple convex sets, and show that these bounds lead to (3). We call any convex set a simple convex set if it can be well approximated by an affine transformation of a rectangle. An extension to simple convex sets is interesting because it allows us to derive similar bounds for $\mathcal{A} = \mathcal{A}^{\mathrm{sp}}(s)$ being the class of (s-)sparsely convex sets. These are sets that can be represented as an intersection of many convex sets whose indicator functions depend nontrivially at most on s elements of their arguments (for some small s).

The sets considered are useful for applications in mathematical statistics. In particular, 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

$$\mathrm{P}\left(\max_{j\in J_k} S_{nj}^X \le t_k \text{ for all } k=1,\dots,\kappa\right) = \mathrm{P}\left(S_n^X \in A\right)$$

can be approximated by $\mathrm{P}(S_n^Y \in A)$ within the error margin $\rho_n(\mathcal{A}^{\mathrm{re}})$; here $\{J_k\}$ are subsets of $\{1,\dots,p\}$, $\{t_k\}$ are thresholds in the interval $(-\infty,\infty)$, $1 \le \kappa < 2^p$ is an integer, and $A \in \mathcal{A}^{\mathrm{re}}$ is a rectangle of the form $\{w \in \mathbb{R}^p : \max_{j\in J_k} w_j \le t_k \text{ for all } k=1,\dots,\kappa\}$. Another example is the probability that a collection of Pearson-type statistics falls below a collection of thresholds

$$\mathrm{P}\Big(\|(S_{nj}^X)_{j\in J_k}\|^2 \le t_k \text{ for all } k=1,\dots,\kappa\Big) = \mathrm{P}\left(S_n^X \in A\right)$$

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

The results in this paper substantially extend those obtained in [15] 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 $J \subset \{1,\dots,p\}$, but in order 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 [9]. Our paper builds upon our previous work [15], which in turn builds on a number of works listed in the bibliography (see [16] for a detailed review and links to the literature).

The organization of this paper is as follows. 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 $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$ 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 $\mathcal{A} = \mathcal{A}^{\mathrm{si}}$ being a class of simple convex sets and for $\mathcal{A} = \mathcal{A}^{\mathrm{sp}}(s)$ being a class of sparsely convex sets. In Section 4, we derive high dimensional Empirical and Multiplier Bootstrap theorems that allow us to approximate $\mathrm{P}(S_n^Y \in A)$ for $A \in \mathcal{A}^{\mathrm{re}}$, $\mathcal{A}^{\mathrm{si}}$, or $\mathcal{A}^{\mathrm{sp}}(s)$ using the data $X_1,\dots,X_n$. In Section 5, we state an induction lemma, a key result underlying the derivations in the paper. Finally, we provide all proofs as well as some technical results in the Appendix.

1.1. 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 $y = (y_1,\dots,y_p)' \in \mathbb{R}^p$, we write $w \le y$ if $w_j \le y_j$ for all $j = 1,\dots,p$. For $y = (y_1,\dots,y_p)' \in \mathbb{R}^p$ and $a \in \mathbb{R}$, we write $y+a = (y_1+a, \dots, y_p+a)'$. 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, $\mathbb{E}_n[x_{ij}] = n^{-1}\sum_{i=1}^n x_{ij}$. We also write $X_1^n := \{X_1, \dots, X_n\}$. For $v \in \mathbb{R}^p$, we use the notation $\|v\|_0 := \sum_{j=1}^p 1\{v_j \neq 0\}$ and $\|v\| = (\sum_{j=1}^p v_j^2)^{1/2}$. 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 real-valued random variable ξ, we define

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

For $\alpha \geq 1$, $\|\cdot\|_{\psi_\alpha}$ 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 $\|\xi_1 + \xi_2\|_{\psi_\alpha} \leq K_\alpha(\|\xi_1\|_{\psi_\alpha} + \|\xi_2\|_{\psi_\alpha})$. Throughout the paper, we assume that $n \geq 4$ and $p \geq 2$.

2. High Dimensional CLT for Rectangles

This section presents a high dimensional CLT for rectangles. We begin with presenting an abstract theorem (Theorem 2.1) that has wide applicability but depends on the tail properties of the distributions of X_{ij}'s in a nontrivial way. Then we apply this theorem under simple moment conditions to derive more explicit bounds in Corollary 2.1.

Let $\mathcal{A}^{\mathrm{re}}$ be the class of all rectangles in $\mathbb{R}^p$; that is, $\mathcal{A}^{\mathrm{re}}$ consists of all sets A of the form

$$A = \{w \in \mathbb{R}^p : a_j \leq w_j \leq b_j \text{ for all } j = 1, \dots, p\} \tag{4}$$

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

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

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

and for $\phi \geq 1$, define

$$M_{n,X}(\phi) := n^{-1}\sum_{i=1}^n \mathrm{E}\left[\max_{1 \leq j \leq p}|X_{ij}|^3 1\left\{\max_{1 \leq j \leq p}|X_{ij}| > \sqrt{n}/(4\phi\log p)\right\}\right]. \tag{5}$$

Similarly, define $M_{n,Y}(\phi)$ with X_{ij}'s replaced by Y_{ij}'s in (5), 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 (Abstract High Dimensional CLT for Rectangles)**.** *Suppose that there exists some constant $b > 0$ such that $n^{-1}\sum_{i=1}^n \mathrm{E}[X_{ij}^2] \geq b$ for all $j = 1, \dots, p$. Then there exist constants $K_1, K_2 > 0$ depending only b such* *that for every constant* $\overline{L}_n \geq L_n$,

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

with

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

Remark 2.1 (Key features of Theorem 2.1)**.** (i) The bound (6) should be contrasted with Bentkus's [7] bound (2). 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 sequence of constants $B_n \geq 1$, $|X_{ij}| \leq B_n$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$. Then it can be shown that the bound (6) reduces to

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \leq K\{n^{-1} B_n^2 \log^7(pn)\}^{1/6} \tag{8}$$

for some universal constant K; see Corollary 2.1 below. Importantly, the right hand side of (8) 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 $\rho_n(\mathcal{A}^{\mathrm{re}}) \to 0$, and if in addition $B_n = O(1)$, the condition reduces to $\log p = o(n^{1/7})$. In contrast, Bentkus's bound (2) requires $p = o(n^{2/7})$ to make $\rho_n(\mathcal{A}) \to 0$ when $\mathcal{A}$ is the class of all convex sets. Hence by restricting the class of sets to the smaller one, $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$, we are able to considerably weaken the requirement on p.

(ii) On the other hand, the bound in (8) 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 that the rate $n^{-1/6}$ is optimal (in a minimax sense) in CLT in infinite dimensional Banach spaces (see [5]), the factor $n^{-1/6}$ seems nearly optimal in terms of dependence on n in the high-dimensional settings as considered here. In addition, examples in [17] suggest that dependence on B_n is also optimal. Hence we conjecture that up to a universal constant,

$$\{n^{-1} B_n^2 (\log p)^a\}^{1/6}$$

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

Remark 2.2 (Relation to previous work)**.** Theorem 2.1 extends Theorem 2.2 in [15] where we derived a bound on $\rho_n(\mathcal{A}^m)$ with $\mathcal{A}^m \subset \mathcal{A}^{\mathrm{re}}$ consisting of all sets of the form

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

for some $a \in \mathbb{R}$. In particular, we improve the dependence on n from $n^{-1/8}$ in [15] to $n^{-1/6}$. In addition, we note that extension to the class $\mathcal{A}^{\mathrm{re}}$ from the class $\mathcal{A}^m$ is not immediate since in both papers we assume that $\mathrm{Var}(S_{nj}^X)$ is bounded below from zero uniformly in $j = 1, \dots, p$, so that it is not possible to directly extend the results in [15] to the class of rectangles $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$ by just rescaling the coordinates in S_n^X. ■

The bound (6) 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 $b, q > 0$ be some constants, 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* $i = 1, \dots, n$,

An application of Theorem 2.1 under these conditions leads to the following corollary. To avoid the repetitions in stating the results below, let

$$D_n^{(1)} = \left(\frac{B_n^2 \log^7(pn)}{n} \right)^{1/6}, \quad D_{n,q}^{(2)} = \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}} \right)^{1/3}. \tag{9}$$

Corollary 2.1 (High Dimensional CLT for Rectangles)**.** *Suppose that conditions* (M.1) *and* (M.2) *are satisfied. Then under* (E.1)*, we have*

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \leq C D_n^{(1)},$$

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

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \leq C\{D_n^{(1)} + D_{n,q}^{(2)}\},$$

where the constant C depends only on b and q.

3. High Dimensional CLT for Simple and Sparsely 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. In addition, it allows us to derive similar bounds for the classes of sparsely convex sets. These classes in turn may be of interest in mathematical statistics where sparse models and techniques have been of canonical importance in the past years.

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

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

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

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

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

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

and 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 $\mathcal{A}^{\text{si}}$ 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 a set A that satisfies condition (C) as a *simple convex set* because it can be well approximated by affine transformations of rectangles. Note that any rectangle $A \in \mathcal{A}^{\text{re}}$ satisfies condition (C) with $a = 0$ and $d = 1$ (recall that $n \geq 4$). Let $A^m(A)$ denote the set A^m appearing in condition (C) applied to the set A.

For every $A \in \mathcal{A}^{\text{si}}$ with an approximating m-generated set $A^m = A^m(A)$ 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 holds:

(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 $\widetilde{X}_1, \dots, \widetilde{X}_n$ rather than to $X_1, \dots, X_n$.

Recall the definition of $\rho_n(\mathcal{A})$ in (1). An extension of Corollary 2.1 leads to the following result in the case where $\mathcal{A} = \mathcal{A}^{\text{si}}$. Recall the definitions of $D_n^{(1)}$ and $D_{n,q}^{(2)}$ given in (9).

Corollary 3.1 (High Dimensional CLT for Simple Convex Sets)**.** *Let $\mathcal{A}^{\text{si}}$ be a class of simple convex sets in $\mathbb{R}^p$ such that conditions* (M.1′), (M.2′), *and* (E.1′) *are satisfied for every $A \in \mathcal{A}^{\text{si}}$. Then*

$$\rho_n(\mathcal{A}^{\text{si}}) \leq C D_n^{(1)}, \tag{10}$$

where the constant C depends only on a, b, and d. If, instead of condition (E.1′)*, condition* (E.2′) *is satisfied for every $A \in \mathcal{A}^{\text{si}}$, then*

$$\rho_n(\mathcal{A}^{\text{si}}) \leq C\{D_n^{(1)} + D_{n,q}^{(2)}\}, \tag{11}$$

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

It is worthwhile to mention that a notable example where the transformed variables $\widetilde{X}_i = (v'X_i)_{v \in \mathcal{V}(A^m)}$ verify condition (E.1$'$) is the case where each X_i obeys a log-concave distribution. Recall that a Borel probability measure ν on $\mathbb{R}^p$ is *log-concave* if for every Borel subsets A_1, A_2 of $\mathbb{R}^p$ and $\lambda \in (0,1)$,

$$\nu(\lambda A_1 + (1-\lambda)A_2) \geq \nu(A_1)^{\lambda}\nu(A_2)^{1-\lambda},$$

where $\lambda A_1 + (1-\lambda)A_2 = \{\lambda x + (1-\lambda)y : x \in A_1, y \in A_2\}$.

Corollary 3.2 (High Dimensional CLT for Simple Convex Sets with Log–concave Distributions)**.** *Suppose that each X_i obeys a centered log-concave distribution on $\mathbb{R}^p$ and that all the eigenvalues of $\mathrm{E}[X_iX_i']$ are bounded from below by a constant $k_1 > 0$ and from above by a constant $k_2 \geq k_1$ for every $i = 1, \dots, n$. Then for $\mathcal{A}^{\mathrm{si}}$ the class of all simple convex sets in $\mathbb{R}^p$, we have*

$$\rho_n(\mathcal{A}^{\mathrm{si}}) \leq Cn^{-1/6}\log^{7/6}(pn),$$

where the constant C depends only on a, b, d, k_1, and k_2.

3.2. **Sparsely convex sets.** We next consider classes of sparsely convex sets defined as follows.

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

Observe that for any s-sparsely convex set $A \subset \mathbb{R}^p$, the integer Q in Definition 3.1 can be chosen to satisfy $Q \leq C_s^p \leq p^s$. Indeed, if we have a sparse representation $A = \cap_{q=1}^Q A_q$ for $Q > C_s^p$, then there are at least two sets A_{q_1} and A_{q_2} with the same main components, and hence we can replace these two sets by one convex set $A_{q_1} \cap A_{q_2}$ with the same main components; this procedure can be repeated until we have $Q \leq C_s^p$.

Example 3.1. The simplest example verifying Definition 3.1 is a rectangle as in (4), which is a 1-sparsely convex set. Another example is the set

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

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 elements of each v_k does not exceed s, this A is an s-sparsely convex set. Yet another example is the set

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

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

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

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 a 2-sparsely convex set. Finally, consider the set

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

where $\{J_k\}$ are subsets of $\{1, \dots, p\}$ of fixed cardinality s, $\{t_k\}$ are thresholds in $(0, \infty)$, and $1 \leq \kappa \leq C_s^p$ is an integer. This A is an s-sparsely convex set.

Fix an integer $s > 0$, and let $\mathcal{A}^{\rm sp}(s)$ denote the class of all s-sparsely convex sets in $\mathbb{R}^p$. We assume that the following condition is satisfied:

(M.1$''$) $n^{-1} \sum_{i=1}^n \mathrm{E}[(v' X_i)^2] \geq b$ *for all* $v \in \mathbb{S}^{p-1}$ *with* $\|v\|_0 \leq s$.

Then we have the following corollary:

Corollary 3.3 (High Dimensional CLT for Sparsely Convex Sets)**.** *Suppose that conditions* (M.1$''$) *and* (M.2) *are satisfied. Then under* (E.1), *we have*

$$\rho_n(\mathcal{A}^{\rm sp}(s)) \leq C D_n^{(1)}, \tag{12}$$

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

$$\rho_n(\mathcal{A}^{\rm sp}(s)) \leq C\{D_n^{(1)} + D_{n,q}^{(2)}\}, \tag{13}$$

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

Remark 3.1 (Dependence on s)**.** In many applications, it may be of interest to consider s-sparsely convex sets with $s = s_n$ depending on n and potentially growing to infinity: $s = s_n \to \infty$. It is therefore interesting to derive the optimal dependence of the constant C in (12) and (13) on s. We leave this question for future work. ■

4. Empirical and Multiplier Bootstrap Theorems

So far we have shown 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 in rectangles $A \in \mathcal{A}^{\rm re}$, simple convex sets $A \in \mathcal{A}^{\rm si}$, or sparsely convex sets $A \in \mathcal{A}^{\rm sp}(s)$. In practice, however, the covariance matrix of S_n^Y is typically unknown, and direct computation of $\mathrm{P}(S_n^Y \in A)$ is infeasible. Therefore, in this section, we derive high dimensional bootstrap theorems which allow us to 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 (for various version of bootstraps, we refer to [30]).

4.1. **Multiplier bootstrap.** We first consider 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

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

in the cases where $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$, $\mathcal{A}^{\mathrm{si}}$, or $\mathcal{A}^{\mathrm{sp}}(s)$.

We begin with the case where $\mathcal{A} = \mathcal{A}^{\mathrm{si}}$. Let

$$\Sigma^{eX} := n^{-1} \sum_{i=1}^n (X_i - \widehat{\mu}_n^X)(X_i - \widehat{\mu}_n^X)', \quad \Sigma^Y := n^{-1} \sum_{i=1}^n \mathrm{E}[Y_i Y_i'].$$

Observe that $\mathrm{E}[S_n^{eX}(S_n^{eX})' \mid X_1^n] = \Sigma^{eX}$ and $\mathrm{E}[S_n^Y(S_n^Y)'] = \Sigma^Y$. Define

$$\Delta_n := \sup_{A \in \mathcal{A}^{\mathrm{si}}} \max_{v_1, v_2 \in \mathcal{V}(A^m(A))} |v_1'(\Sigma^{eX} - \Sigma^Y)v_2|.$$

Then we have the following theorem for classes of simple convex sets.

Theorem 4.1 (Abstract Multiplier Bootstrap Theorem for Simple Convex Sets)**.** *Suppose that condition* (M.1$'$) *is satisfied for every $A \in \mathcal{A}^{\mathrm{si}}$. Then for every constant $\bar{\Delta}_n > 0$, on the event $\Delta_n \leq \bar{\Delta}_n$, we have*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{si}}) \leq C \left\{ \bar{\Delta}_n^{1/3} \log^{2/3}(pn) + n^{-1} \log^{1/2}(pn) \right\},$$

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

Remark 4.1 (Case of rectangles)**.** From the proof of Theorem 4.1, we have the following bound when $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$: under (M.1), for every constant $\bar{\Delta}_n > 0$, on the event $\Delta_{n,r} \leq \bar{\Delta}_n$, we have

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{re}}) \leq C \bar{\Delta}_n^{1/3} \log^{2/3} p,$$

where the constant C depends only on b, and $\Delta_{n,r}$ is defined by

$$\Delta_{n,r} = \max_{1 \leq j,k \leq p} |\Sigma_{jk}^{eX} - \Sigma_{jk}^Y|,$$

where Σ_{jk}^{eX} and Σ_{jk}^Y are the (j,k)-th elements of Σ^{eX} and Σ^Y, respectively. ■

We shall derive more explicit bounds on $\rho_n^{MB}(\mathcal{A}^{\mathrm{si}})$ under suitable moment conditions as in the previous section. We will need to strengthen condition (C) and will assume that all sets A in $\mathcal{A}^{\mathrm{si}}$ 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 $m \leq (pn)^d$ and A^m is such that for $v \in \mathcal{V}(A^m)$, $\|v\|_0 \leq s$.*

Note that condition (C$'$) is more restrictive than (C) as it requires that the outward unit normal vectors to the hyperplanes forming the m-generated convex set A^m are sparse. We need this extra condition to control Δ_n. Then we have the following corollary. Here for $\alpha \in (0,1)$, define

$$D_n^{(1)}(\alpha) = \left(\frac{B_n^2 (\log^5(pn)) \log^2(1/\alpha)}{n} \right)^{1/6}, \ D_{n,q}^{(2)}(\alpha) = \left(\frac{B_n^2 \log^3 p}{\alpha^{2/q} n^{1-2/q}} \right)^{1/3}.$$

Corollary 4.1 (Multiplier Bootstrap for Simple Convex Sets). *Let $\alpha \in (0, e^{-1})$ be a constant. Suppose that conditions* (C′) *and* (M.1′) *are satisfied for every $A \in \mathcal{A}^{\mathrm{si}}$. In addition, suppose that condition* (M.2) *is satisfied. Then under* (E.1)*, we have with probability at least $1-\alpha$,*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{si}}) \leq C D_n^{(1)}(\alpha),$$

where the constant C depends only on a, b, d and s; while under (E.2)*, we have with probability at least $1-\alpha$,*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{si}}) \leq C\{D_n^{(1)}(\alpha) + D_{n,q}^{(1)}(\alpha)\},$$

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

When each X_i obeys a log-concave distribution, then we have the following corollary analogous to Corollary 3.2. In this case, in stead of condition (C′), we will make an alternative assumption that the cardinality of $\cup_{A \in \mathcal{A}^{\mathrm{si}}} \mathcal{V}(A^m(A))$ for each A is at most $(pn)^d$.

Corollary 4.2 (Multiplier Bootstrap for Simple Convex Sets with Log-concave Distributions). *Let $\alpha \in (0, e^{-1})$ be a constant. Suppose that each X_i obeys a centered log-concave distribution on $\mathbb{R}^p$ and that all the eigenvalues of* $\mathrm{E}[X_i X_i']$ *are bounded from below by a constant $k_1 > 0$ and from above by a constant $k_2 \geq k_1$ for all $i = 1, \dots, n$. Moreover, suppose that every $A \in \mathcal{A}^{\mathrm{si}}$ satisfies, in addition to condition (C), that the cardinality of the set $\cup_{A \in \mathcal{A}^{\mathrm{si}}} \mathcal{V}(A^m(A))$ is at most $(pn)^d$. Then with probability at least $1-\alpha$,*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{si}}) \leq C n^{-1/6} (\log^{5/6}(pn)) \log^{1/3}(1/\alpha),$$

where the constant C depends only on a, d, k_1, and k_2.

Finally we shall derive explicit bounds on $\rho_n^{MB}(\mathcal{A})$ in the case where $\mathcal{A}$ is the class of all s-sparsely convex sets: $\mathcal{A} = \mathcal{A}^{\mathrm{sp}}(s)$.

Corollary 4.3 (Multiplier Bootstrap for Sparsely Convex Sets). *Let $\alpha \in (0, e^{-1})$ be a constant. Suppose that conditions* (M.1″) *and* (M.2) *are satisfied. Then under* (E.1)*, we have with probability at least $1-\alpha$,*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{sp}}(s)) \leq C D_n^{(1)}(\alpha), \tag{14}$$

where the constant C depends only on b and s; while under (E.2)*, we have with probability at least $1-\alpha$,*

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{sp}}(s)) \leq C\{D_n^{(1)}(\alpha) + D_{n,q}^{(2)}(\alpha)\}, \tag{15}$$

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

4.2. **Empirical bootstrap.** Here we consider the empirical bootstrap. For brevity, we shall focus here on the cases where $\mathcal{A}$ is $\mathcal{A}^{\mathrm{re}}$ or $\mathcal{A}^{\mathrm{si}}$. 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

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

in the cases where $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$ or $\mathcal{A}^{\mathrm{si}}$. To state the bound, define

$$\widehat{L}_n := \max_{1 \le j \le p} \sum_{i=1}^n |X_{ij} - \widehat{\mu}_{nj}^X|^3 / n,$$

which is an empirical analog of L_n, and for $\phi \geq 1$, define

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

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

which are empirical analogs of $M_{n,X}(\phi)$ and $M_{n,Y}(\phi)$, respectively. 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 $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$.

Theorem 4.2 (Abstract Empirical Bootstrap Theorem)**.** *For arbitrary positive constants b, $\overline{L}_n$, and $\overline{M}_n$, the inequality*

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

holds on the event

$$\{\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj}^X)^2] \ge b \text{ for all } j = 1, \dots, p\} \cap \{\widehat{L}_n \le \overline{L}_n\} \cap \{\widehat{M}_n \le \overline{M}_n\},$$

where

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

Here $K_1, K_2 > 0$ are constants that depend only on b.

As in the multiplier bootstrap, we shall derive explicit bounds on $\rho_n^{EB}(\mathcal{A})$ under suitable moment conditions. Here we only state the results for classes of simple convex sets $\mathcal{A} = \mathcal{A}^{\mathrm{si}}$ but note that the same result applies to the case of rectangles since a rectangle is a special case of a simple convex set.

Corollary 4.4 (Empirical Bootstrap for Simple Convex Sets). *Let* $\alpha \in (0, e^{-1})$ *be a constant and suppose that* $\log(1/\alpha) \le K \log(pn)$ *for some other constant* K. *Moreover, suppose that all the assumptions in Corollary 4.1 except for* (E.1) *and* (E.2) *are satisfied. Then under* (E.1)*, we have with probability at least* $1 - \alpha$*,*

$$\rho_n^{EB}(\mathcal{A}^{\mathrm{si}}) \le C D_n^{(1)}, \tag{16}$$

where the constant C *depends only on* $a, b, d, s,$ *and* K*; while under* (E.2)*, we have with probability at least* $1 - \alpha$*,*

$$\rho_n^{EB}(\mathcal{A}^{\mathrm{si}}) \le C\{D_n^{(1)} + D_{n,q}^{(2)}(\alpha)\}, \tag{17}$$

where the constant C *depends only on* $a, b, d, s, q,$ *and* K.

When each X_i obeys a log-concave distribution, then we have the following corollary.

Corollary 4.5 (Empirical Bootstrap for Simple Convex Sets with Log–concave Distributions). *Let* $\alpha \in (0, e^{-1})$ *be a constant and suppose that* $\log(1/\alpha) \le K \log(pn)$ *for some other constant* K. *Moreover, suppose that all the assumptions in Corollary 4.2 are satisfied. Then with probability at least* $1 - \alpha$*,*

$$\rho_n^{EB}(\mathcal{A}^{\mathrm{si}}) \le C n^{-1/6} \log^{7/6}(pn),$$

where the constant C *depends only on* $a, d, k_1, k_2,$ *and* K.

Remark 4.2 (Bootstrap CLTs in a.s. sense). Corollaries 4.1 and 4.4 lead to the following multiplier and empirical bootstrap CLTs in the a.s. sense. Suppose that all the assumptions in Corollary 4.1 except for (E.1) and (E.2) are satisfied. We allow $p = p_n \to \infty$ and $B_n \to \infty$ as $n \to \infty$ but assume that a, b, d, q, s are all fixed. Then by applying Corollaries 4.1 and 4.4 with $\alpha = \alpha_n = n^{-1} \log^{-2} n$, together with the Borel-Cantelli lemma (note that $\sum_{n=4}^{\infty} n^{-1} \log^{-2} n < \infty$), we have with probability one

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{si}}) \vee \rho_n^{EB}(\mathcal{A}^{\mathrm{si}}) = \begin{cases} O\{D_n^{(1)}\} & \text{under (E.1)} \\ O\{D_n^{(1)} \vee D_{n,q}^{(2)}(\alpha_n)\} & \text{under (E.2),} \end{cases}$$

and it is routine to verify that $D_n^{(1)} = o(1)$ if $B_n^2 \log^7(pn) = o(n)$, and $D_{n,q}^{(2)}(\alpha_n) = o(1)$ if $B_n^2(\log^3(pn)) \log^{4/q} n = o(n^{1-4/q})$.

5. Induction Lemma

In this section, we state a lemma that plays a key role in the proof of our high dimensional CLT for rectangles (Theorem 2.1). Define

$$\varrho_n := \sup_{y \in \mathbb{R}^p, v \in [0,1]} |\mathrm{P}\left(\sqrt{v} S_n^X + \sqrt{1-v} S_n^Y \le y\right) - \mathrm{P}(S_n^Y \le y)|,$$

and recall that $M_n(\phi) := M_{n,X}(\phi) + M_{n,Y}(\phi)$ for $\phi \ge 1$. The lemma below provides a bound on ϱ_n.

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

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

up to a constant K that depends only on b.

Lemma 5.1 has an immediate corollary. Indeed, define

$$\varrho_n' := \sup_{A \in \mathcal{A}^{\mathrm{re}}, v \in [0,1]} |\mathrm{P}(\sqrt{v} S_n^X + \sqrt{1-v} S_n^Y \in A) - \mathrm{P}(S_n^Y \in A)|$$

where $\mathcal{A}^{\mathrm{re}}$ 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 $n^{-1}\sum_{i=1}^{n} \mathrm{E}[X_{ij}^2] \geq b$ for all $j = 1, \dots, p$. Then ϱ_n' satisfies the following inequality for all $\phi \geq 1$:*

$$\varrho_n' \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}} \Big\{\phi L_n \varrho_n' + L_n \log^{1/2} p + \phi M_n(2\phi)\Big\} + \frac{\log^{1/2} p}{\phi}$$

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

Appendix A. Anti-concentration inequalities

One of the main ingredients of the proof of Lemma 5.1 (and the proofs of the other results indeed) is the following anti-concentration due to [28].

Lemma A.1 (Nazarov's inequality, [28])**.** *Let $Y = (Y_1, \dots, Y_p)'$ be a centered Gaussian random vector in $\mathbb{R}^p$ such that $\mathrm{E}[Y_j^2] \geq b$ for all $j = 1, \dots, p$ and some constant $b > 0$. Then for every $y \in \mathbb{R}^p$ and $a > 0$,*

$$\mathrm{P}(Y \leq y + a) - \mathrm{P}(Y \leq y) \leq Ca(\log p)^{1/2},$$

where C is a constant depending only on b.

Remark A.1. This inequality is less sharp than the dimension-free anti-concentration bound $Ca\mathrm{E}[\max_{1\leq j\leq p} Y_j]$ proved in [18] for the case of max rectangles. 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 $\mathrm{E}[\max_{1\leq j\leq p} Y_j]$. This also makes the anti-concentration bound in [18] 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 [18] for classes of rectangles other than max rectangles. ■

Proof of Lemma A.1. Let $\Sigma = \mathrm{E}[YY']$; then Y has the same distribution as $\Sigma^{1/2} Z$ where Z is a standard Gaussian random vector. Write $\Sigma^{1/2} =$

$(\sigma_1, \dots, \sigma_p)'$ where each σ_j is a p-dimensional vector. Note that $\|\sigma_j\| = (\mathrm{E}[Y_j^2])^{1/2} \geq b^{1/2}$. Then

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

and similarly

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

Since Z is a standard Gaussian random vector, and $a/\|\sigma_j\| \leq a/b^{1/2}$ for all $j = 1, \dots, p$, the assertion follows from Theorem 20 in [23], whose proof the authors credit to Nazarov [28]. ■

We will use another anti-concentration inequality by [28] in the proofs for Section 3, which is an extension of Theorem 4 in [3].

Lemma A.2. *Let A be a $p \times p$ symmetric positive definite matrix, and let $\gamma_A = N(0, A^{-1})$. Then there exists a universal constant $C > 0$ such that for every convex set $Q \subset \mathbb{R}^p$,*

$$\limsup_{h \downarrow 0} \frac{\gamma_A(Q^h \setminus Q)}{h} \leq C \sqrt{\|A\|_{HS}},$$

where $\|A\|_{HS}$ is the Hilbert-Schmidt norm of A.

Proof. See [28]. ■

Appendix B. Proof of Lemma 5

We begin with stating the following variants of Chebyshev's association inequality.

Lemma B.1. *Let $\varphi_i : \mathbb{R} \to [0, \infty)$, $i = 1, 2$ be non-increasing functions, and let $\xi_i, i = 1, 2$ be independent real-valued random variables. Then*

$$\mathrm{E}[\varphi_1(\xi_1)]\mathrm{E}[\varphi_2(\xi_1)] \leq \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_1)], \tag{18}$$

$$\mathrm{E}[\varphi_1(\xi_1)]\mathrm{E}[\varphi_2(\xi_2)] \leq \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_1)] + \mathrm{E}[\varphi_1(\xi_2)\varphi_2(\xi_2)], \tag{19}$$

$$\mathrm{E}[\varphi(\xi_1)\varphi_2(\xi_2)] \leq \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_1)] + \mathrm{E}[\varphi_1(\xi_2)\varphi_2(\xi_2)], \tag{20}$$

where we assume that all the expectations exist and are finite. Moreover, (20) holds without independence of ξ_1 and ξ_2.

Proof of Lemma B.1. The inequality (18) is Chebyshev's association inequality; see Theorem 2.14 in [11]. Moreover, since ξ_1 and ξ_2 are independent, (19) follows from (20). In turn, (20) follows from

$$\begin{aligned} \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_2)] &\leq \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_2)] + \mathrm{E}[\varphi_2(\xi_1)\varphi_1(\xi_2)] \\ &\leq \mathrm{E}[\varphi_1(\xi_1)\varphi_2(\xi_1)] + \mathrm{E}[\varphi_1(\xi_2)\varphi_2(\xi_2)], \end{aligned}$$

where the first inequality follows from the fact that $\varphi_2(\xi_1)\varphi_1(\xi_2) \geq 0$, and the second inequality follows from rearranging the terms in the following inequality:

$$\mathrm{E}[(\varphi_1(\xi_1) - \varphi_1(\xi_2))(\varphi_2(\xi_1) - \varphi_2(\xi_2))] \geq 0,$$

which follows from monotonicity of φ_1 and φ_2. ■

Proof of Lemma 5.1. The proof relies on a Slepian-Stein method developed in [15]. Here the notation $\lesssim$ means that the left-hand side is bounded by the right hand side up to some constant depending only on b.

We begin with preparing some notation. Let $W_1, \dots, W_n$ be a copy of $Y_1, \dots, Y_n$. Without loss of generality, we may assume that $X_1, \dots, X_n$, $Y_1, \dots, Y_n$, and $W_1, \dots, W_n$ are independent. Consider

$$S_n^W := \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i.$$

Then $\mathrm{P}(S_n^Y \leq y) = \mathrm{P}(S_n^W \leq y)$, so that

$$\varrho_n = \sup_{s \in \mathbb{R}^p, v \in [0,1]} |\mathrm{P}\left(\sqrt{v} S_n^X + \sqrt{1-v} S_n^Y \leq y\right) - \mathrm{P}(S_n^W \leq y)|.$$

Pick any $y \in \mathbb{R}^p$ and $v \in [0,1]$. Let $\beta := \phi \log p$, and define the function

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

The function $F_\beta(w)$ has the following property:

$$0 \leq F_\beta(w) - \max_{1 \leq j \leq p} (w_j - y_j) \leq \beta^{-1} \log p = \phi^{-1}, \text{ for all } w \in \mathbb{R}^p. \tag{21}$$

Consider a thrice continuously differentiable function $g_0 : \mathbb{R} \to [0,1]$ whose derivatives up to the third order are all bounded such that $g_0(t) = 1$ for $t \leq 0$ and $g_0(t) = 0$ for $t \geq 1$. Define $g(t) := g_0(\phi t), t \in \mathbb{R}$, and

$$m(w) := g(F_\beta(w)), \ w \in \mathbb{R}^p.$$

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 [15]: for every $j, k, l = 1, \dots, p$, there exists a function $U_{jkl}(w)$ such that

$$|m_{jkl}(w)| \leq U_{jkl}(w), \tag{22}$$

$$\textstyle\sum_{j,k,l=1}^p U_{jkl}(w) \lesssim (\phi^3 + \phi\beta + \phi\beta^2) \lesssim \phi\beta^2, \tag{23}$$

$$U_{jkl}(w) \lesssim U_{jkl}(w + \widetilde{w}) \lesssim U_{jkl}(w), \tag{24}$$

where the inequalities (22) and (23) hold for all $w \in \mathbb{R}^p$, and the inequality (24) holds for all $w, \widetilde{w} \in \mathbb{R}^p$ with $\max_{1 \leq j \leq p} |\widetilde{w}_j| \beta \leq 1$ (formally, [15] only considered the case where $y = (0, \dots, 0)'$ but the extension to $y \in \mathbb{R}^p$ is trivial). Moreover, define the functions

$$h(w,t) := 1\left\{-\phi^{-1} - t/\beta < \max_{1 \le j \le p}(w_j - y_j) \le \phi^{-1} + t/\beta\right\},\ w \in \mathbb{R}^p, t > 0, \tag{25}$$

$$\omega(t) := \frac{1}{\sqrt{t} \wedge \sqrt{1-t}},\ t \in (0,1).$$

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

$$|\mathrm{E}[\mathcal{I}_n]| \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}}\left(\phi L_n \varrho_n + L_n \log^{1/2} p + \phi M_n(\phi)\right) \tag{26}$$

where

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

The second step combines this bound with Lemma A.1 to complete the proof.

Step 1. Define the Slepian interpolant

$$Z(t) := \sum_{i=1}^n Z_i(t),\ t \in [0,1],$$

where

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

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

$$\mathcal{I}_n = m(\sqrt{v} S_n^X + \sqrt{1-v} S_n^Y) - m(S_n^W) = \int_0^1 \frac{dm(Z(t))}{dt} dt. \tag{27}$$

Denote by $Z^{(i)}(t)$ the Stein leave-one-out term for $Z(t)$:

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

Finally, define

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

For brevity of notation, we omit the 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, from (27) and Taylor's theorem, we have

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

where

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

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

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

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

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

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

To this end, let

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

and decompose III as $III = III_1 + III_2$, where

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

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

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

$$\begin{aligned} |III_2| &\le \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \int_0^1 \mathrm{E}[(1-\chi_i)U_{jkl}(Z^{(i)}+\tau Z_i)|\dot{Z}_{ij}Z_{ik}Z_{il}|]d\tau dt \\ &\lesssim \phi\beta^2 \sum_{i=1}^{n} \int_0^1 \mathrm{E}[(1-\chi_i)\max_{1\le j,k,l\le p}|\dot{Z}_{ij}Z_{ik}Z_{il}|]dt \\ &\lesssim \frac{\phi\beta^2}{n^{3/2}} \sum_{i=1}^{n} \int_0^1 \omega(t)\mathrm{E}[(1-\chi_i)\max_{1\le j\le p}|X_{ij}|^3 \vee |Y_{ij}|^3 \vee |W_{ij}|^3]dt, \end{aligned} \quad (28)$$

where the first and the second inequalities follow from (22) and (23), respectively. Moreover, by letting $\mathcal{T} = \sqrt{n}/(4\beta)$ and using the union bound, we have

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

Hence, using the inequality

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

together with the inequality (20) in Lemma B.1, we conclude that the integral in (28) is bounded from above up to a universal constant by

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

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

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

To bound III_1, recall the definition of $h(w,t)$ in (25). Note that $m_{jkl}(Z^{(i)}+\tau Z_i)=0$ for all $\tau\in[0,1]$ whenever $h(Z^{(i)},2)=0$ and $\chi_i=1$. Indeed if $\chi_i=1$, then $\max_{1\le j\le p}|Z_{ij}|\le 3/(4\beta)\le 1/\beta$, and so when $h(Z^{(i)},2)=0$ and $\chi_i=1$, we have $h(Z^{(i)}+\tau Z_i,0)=0$, which in turn implies that either $F_\beta(Z^{(i)}+\tau Z_i)\le 0$ or $F_\beta(Z^{(i)}+\tau Z_i)\ge\phi^{-1}$ because of (21); in both cases, the assertion follows from the definitions of m and g. Hence

$$\begin{aligned}|III_1|&\le\sum_{j,k,l=1}^p\sum_{i=1}^n\int_0^1\int_0^1\mathrm{E}[\chi_i|m_{jkl}(Z^{(i)}+\tau Z_i)\dot Z_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^p\sum_{i=1}^n\int_0^1\int_0^1\mathrm{E}[\chi_i h(Z^{(i)},2)U_{jkl}(Z^{(i)}+\tau Z_i)|\dot Z_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^p\sum_{i=1}^n\int_0^1\int_0^1\mathrm{E}[\chi_i h(Z^{(i)},2)U_{jkl}(Z^{(i)})|\dot Z_{ij}Z_{ik}Z_{il}|]d\tau dt\\&\lesssim\sum_{j,k,l=1}^p\sum_{i=1}^n\int_0^1\mathrm{E}[h(Z^{(i)},2)U_{jkl}(Z^{(i)})]\mathrm{E}[|\dot Z_{ij}Z_{ik}Z_{il}|]dt,\end{aligned}\tag{29}$$

where the second inequality follows from (22), the third inequality from (24), and the fourth inequality from the indepence of $Z^{(i)}$ from $\dot Z_{ij}Z_{ik}Z_{il}$. Then we split the integral in (29) by inserting $\chi_i+(1-\chi_i)$ under the first expectation sign. We have

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

where the last inequality follows from the argument similar to that used to bound III_2 with applying (18) and (19) instead of (20) in Lemma B.1. Moreover, since $h(Z^{(i)},2)=0$ whenever $h(Z,4)=0$ and $\chi_i=1$ (which follows from the same argument as before), we have

$$
\begin{aligned}
&\sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}[\chi_i h(Z^{(i)}, 2) U_{jkl}(Z^{(i)})] \mathrm{E}[|\dot{Z}_{ij} Z_{ik} Z_{il}|] dt \\
&\quad \lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}[h(Z,4) U_{jkl}(Z)] \mathrm{E}[|\dot{Z}_{ij} Z_{ik} Z_{il}|] dt \\
&\quad = \sum_{j,k,l=1}^{p} \int_0^1 \mathrm{E}[h(Z,4) U_{jkl}(Z)] \sum_{i=1}^{n} \mathrm{E}[|\dot{Z}_{ij} Z_{ik} Z_{il}|] dt \\
&\quad \lesssim \phi\beta^2 \int_0^1 \mathrm{E}[h(Z,4)] \max_{1 \le j,k,l \le p} \sum_{i=1}^{n} \mathrm{E}[|\dot{Z}_{ij} Z_{ik} Z_{il}|] dt. \qquad (30)
\end{aligned}
$$

To bound (30), observe that

$$
|\dot{Z}_{ij} Z_{ik} Z_{il}| \lesssim \frac{\omega(t)}{n^{3/2}} (|X_{ij}|^3 + |Y_{ij}|^3 + |W_{ij}|^3),
$$

which, together with the facts that $\mathrm{E}[|W_{ij}|^3] = \mathrm{E}[|Y_{ij}|^3]$ and $\mathrm{E}[|Y_{ij}|^3] \lesssim (\mathrm{E}[|Y_{ij}|^2])^{3/2} = (\mathrm{E}[|X_{ij}|^2])^{3/2} \le \mathrm{E}[|X_{ij}|^3]$, implies that

$$
\max_{1 \le j,k,l \le p} \sum_{i=1}^{n} \mathrm{E}[|\dot{Z}_{ij} Z_{ik} Z_{il}|] \lesssim \frac{\omega(t)}{n^{3/2}} \max_{1 \le j,k \le p} \sum_{i=1}^{n} (\mathrm{E}[|X_{ij}|^3] + \mathrm{E}[|Y_{ij}|^3]) \lesssim \frac{\omega(t)}{n^{1/2}} L_n.
$$

Meanwhile, observe that

$$
\mathrm{E}[h(Z,4)] = \mathrm{P}(\widetilde{V}_n \le \overline{I}) - \mathrm{P}(\widetilde{V}_n \le \underline{I}),
$$

where

$$
\begin{aligned}
\widetilde{V}_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sqrt{tv} X_i + \sqrt{t(1-v)} Y_i + \sqrt{1-t} W_i) \\
&\stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sqrt{tv} X_i + \sqrt{1-tv} Y_i),
\end{aligned}
$$

and $\underline{I} = y - \phi^{-1} - 4\beta^{-1}$, $\overline{I} = y + \phi^{-1} + 4\beta^{-1}$; here the notation $\stackrel{d}{=}$ denotes equality in distribution, and $\underline{I}$ and $\overline{I}$ are vectors in $\mathbb{R}^p$ (recall the rules of summation of vectors and scalars defined in Section 1.1). Now by the definition of ϱ_n,

$$
\mathrm{P}(\widetilde{V}_n \le \overline{I}) \le \mathrm{P}(S_n^Y \le \overline{I}) + \varrho_n, \quad \mathrm{P}(\widetilde{V}_n \le \underline{I}) \ge \mathrm{P}(S_n^Y \le \underline{I}) - \varrho_n,
$$

and by Lemma A.1,

$$
\mathrm{P}(S_n^Y \le \overline{I}) - \mathrm{P}(S_n^Y \le \underline{I}) \lesssim \phi^{-1} \log^{1/2} p
$$

since $\beta^{-1} \lesssim \phi^{-1}$ and $\mathrm{E}[(S_{nj}^Y)^2] = \mathrm{E}[(S_{nj}^X)^2] = n^{-1} \sum_{i=1}^{n} \mathrm{E}[X_{ij}^2] \ge b$ for all $j = 1, \dots, p$. Hence

$$
\mathrm{E}[h(Z,4)] \lesssim \varrho_n + \phi^{-1} \log^{1/2} p.
$$

By these bounds, together with the fact that $\int_0^1 \omega(t)dt \lesssim 1$, we conclude that

$$(30) \lesssim \frac{\phi\beta^2 L_n}{n^{1/2}}(\varrho_n + \phi^{-1}\log^{1/2} p) \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}}(\phi L_n \varrho_n + L_n \log^{1/2} p),$$

where we have used $\beta = \phi \log p$. The desired assertion (26) then follows.

Step 2. We are now in position to finish the proof. Let

$$V_n := \sqrt{v} S_n^X + \sqrt{1-v} S_n^Y.$$

Then we have

$$\begin{aligned}
\mathrm{P}(V_n \leq y - \phi^{-1}) &\leq \mathrm{P}(F_\beta(V_n) \leq 0) \leq \mathrm{E}[m(V_n)] \\
&\leq \mathrm{P}(F_\beta(S_n^W) \leq \phi^{-1}) + (\mathrm{E}[m(V_n)] - \mathrm{E}[m(S_n^W)]) \\
&\leq \mathrm{P}(S_n^W \leq y + \phi^{-1}) + \mathrm{E}[\mathcal{I}_n] \\
&\leq \mathrm{P}(S_n^W \leq y - \phi^{-1}) + C\phi^{-1}\log^{1/2} p + \mathrm{E}[\mathcal{I}_n],
\end{aligned}$$

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 last inequality follows from Lemma A.1. Here the constant C depends only on b. Likewise we have

$$\mathrm{P}(V_n \leq y - \phi^{-1}) \geq \mathrm{P}(S_n^W \leq y - \phi^{-1}) - C\phi^{-1}\log^{1/2} p + \mathrm{E}[\mathcal{I}_n].$$

The conclusion of the lemma follows from combining these inequalities with the bound on $|\mathrm{E}[\mathcal{I}_n]|$ derived in Step 1. ■

Proof of Corollary 5.1. Pick any rectangle

$$A = \{w \in \mathbb{R}^p : w_j \in [a_j, b_j] \text{ for all } j = 1, \dots, p\}.$$

For $i = 1, \dots, n$, consider the 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 $j = p+1, \dots, 2p$. Then

$$\mathrm{P}(S_n^X \in A) = \mathrm{P}(S_n^{\widetilde{X}} \leq y), \quad \mathrm{P}(S_n^Y \in A) = \mathrm{P}(S_n^{\widetilde{Y}} \leq y),$$

where the vector $y \in \mathbb{R}^{2p}$ is defined by $y_j = b_j$ for $j = 1, \dots, p$ and $y_j = -a_{j-p}$ for $j = p+1, \dots, 2p$, and $S_n^{\widetilde{X}}$ and $S_n^{\widetilde{Y}}$ are defined as S_n^X and S_n^Y with X_i's and Y_i's replaced by $\widetilde{X}_i$'s and $\widetilde{Y}_i$'s. Hence the corollary follows from applying Lemma 5.1 to $\widetilde{X}_1, \dots, \widetilde{X}_n$ and $\widetilde{Y}_1, \dots, \widetilde{Y}_n$. ■

Appendix C. Proofs for Section 2

Proof of Theorem 2.1. The proof relies on Lemma 5.1 and its Corollary 5.1. 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 (7), so that

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

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

Then applying Corollary 5.1 with $\phi = \phi_n/2$, we have

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

Since $8(K' \vee 1)^2 > 1$, solving this inequality for ϱ_n' and observing that $\rho_n(\mathcal{A}^{\mathrm{re}}) \leq \varrho_n'$ leads to the desired assertion. ∎

Before proving Corollary 2.1, we shall verify the following elementary inequality.

Lemma C.1. *Let ξ be a non-negative random variable such that* $\mathrm{P}(\xi > x) \leq A e^{-x/B}$ *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 of Lemma C.1. 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 $\mathrm{P}(\xi > x) \leq A e^{-x/B}$, 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

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

∎

Proof of Corollary 2.1. The proof relies on application of Theorem 2.1. Without loss of generality, we may assume that

$$\{D_n^{(1)}\}^6 = \frac{B_n^2 \log^7(pn)}{n} \leq c := \min\{(c_1/2)^3, (K_2/2)^6\}, \tag{31}$$

where K_2 appears in (7) and $c_1 > 0$ is a constant that depends only on b (c_1 will be defined later), since otherwise we can make the assertions trivial by setting C large enough.

Now by Theorem 2.1, we have

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

where $\phi_n = K_2\{n^{-1}\overline{L}_n^2 \log^4 p\}^{-1/6}$, and $\overline{L}_n$ is any constant such that $\overline{L}_n \geq L_n$. Recall that

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

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

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

It remains to choose a suitable constant $\overline{L}_n$ such that $\overline{L}_n \geq L_n$ and bound $M_{n,X}(\phi_n)$ and $M_{n,Y}(\phi_n)$. To this end, we consider cases (E.1) and (E.2) separately. In what follows, the notation $\lesssim$ means that the left hand side is bounded by the right hand side up to a positive constant that depends only on b under case (E.1), and on b and q under case (E.2).

Case (E.1). 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. Hence by Lemma 2.2.2 in [40], we have for some universal constant $C_1 > 0$, $\|\max_{1\leq j\leq p} X_{ij}\|_{\psi_1} \leq C_1 B_n \log p$, which, together with Markov's inequality, implies that for every $t > 0$,

$$\mathrm{P}\left(\max_{1\leq j\leq p}|X_{ij}| > t\right) \leq 2\exp\left(-\frac{t}{C_1 B_n \log p}\right).$$

Applying Lemma C.1, we have

$$M_{n,X}(\phi_n) \lesssim (\sqrt{n}/(\phi_n\log p) + B_n\log p)^3 \exp\left(-\frac{\sqrt{n}}{4C_1\phi_n B_n\log^2 p}\right).$$

Here

$$\begin{aligned}\frac{\sqrt{n}}{4C_1\phi_n B_n \log^2 p} &= \frac{c_1 n^{1/3}}{B_n^{2/3}\log^{4/3} p} \quad \left(c_1 := \frac{1}{4K_2C_1}\right)\\ &\geq c_1 c^{-1/3}\log(pn) \geq 2\log(pn). \quad \text{(by (31))}.\end{aligned}$$

Moreover, by (31) and $\phi_n^{-1} = K_2^{-1}\{n^{-1}B_n^2\log^4 p\}^{1/6} \leq c^{1/6}/K_2 \leq 1$, we have $(\sqrt{n}/(\phi_n\log p) + B_n\log p)^3 \lesssim n^{2/3}$, which implies that

$$M_{n,X}(\phi_n) \lesssim n^{2/3}\exp(-2\log(pn)) \leq n^{-1/2}.$$

For $M_{n,Y}(\phi_n)$, since $\mathrm{E}[Y_{ij}^2] = \mathrm{E}[X_{ij}^2] \leq C_1 B_n^2$ and hence $\|Y_{ij}\|_{\psi_1} \lesssim B_n$ for all i and j (as each Y_{ij} is Gaussian), we also have $M_{n,Y}(\phi_n) \lesssim n^{-1/2}$. The conclusion of the corollary in this case follows from the fact that $n^{-1/2}B_n^{-1} \leq D_n^{(1)}$.

Case (E.2). Without loss of generality, in addition to (31), we may assume that

$$\{D_{n,q}^{(2)}\}^{3/2} = \frac{B_n\log^{3/2} p}{n^{1/2-1/q}} \leq (K_2/2)^{3/2}. \tag{32}$$

We begin with noting that

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

As the map $x \mapsto x^{1/6}$ is sub-linear, $\{n^{-1}\overline{L}_n^2 \log^7 p\}^{1/6} \le D_n^{(1)} + D_{n,q}^{(2)} \le K_2$, so that by (31) and (32), $\phi_n^{-1} = K_2^{-1}\{n^{-1}\overline{L}_n^2 \log^4 p\}^{1/6} \le c^{1/6}/K_2 \le 1$.

Note that for any real-valued random variable Z and any $t > 0$, $\mathrm{E}[|Z|^3 1(|Z| > t)] \le \mathrm{E}[|Z|^3 (|Z|/t)^{q-3} 1(|Z| > t)] \le t^{3-q}\mathrm{E}[|Z|^q]$. Hence

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

Here using the bound $\overline{L}_n^{-1} \le B_n^{-2} n^{1/2-2/q} \log^{1/2} p$, we have that $\phi_n \lesssim n^{1/3-2/(3q)} B_n^{-2/3} \log^{-1/2} p$, so that

$$M_{n,X}(\phi_n) \lesssim \frac{B_n^{q/3+2} (\log p)^{q/2-3/2}}{n^{q/6+1/6-2/q}},$$

which implies that $M_{n,X}(\phi_n)/\overline{L}_n \lesssim D_{n,q}^{(2)}$. Meanwhile, as in the previous case, we have $M_{n,Y}(\phi_n) \lesssim n^{-1/2}$, which leads to the desired conclusion in this case. ∎

Appendix D. Proofs for Section 3

Proof of Corollary 3.1. Here C denotes a positive constant that depends only on a, b, and d if (E.1′) is satisfied, and on a, b, d, and q if (E.2′) is satisfied; the value of C may change from place to place. Pick any $A \in \mathcal{A}^{\mathrm{si}}$. Let A^m be an approximating m-generated convex set as in (C.1). By assumption, $A^m \subset A \subset A^{m,\epsilon}$, so that by letting

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

we have

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

Likewise we have $\mathrm{P}(S_n^X \in A) \ge \mathrm{P}(S_n^Y \in A) - C\epsilon \log^{1/2} p - \overline{\rho}$, by which we conclude

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

Recalling that $\epsilon = a/n$ and $B_n \ge 1$, we have $\epsilon \log^{1/2} p \le C D_n^{(1)}$. Hence the assertions of the corollary follow if we prove

$$\overline{\rho} \le \begin{cases} C D_n^{(1)} & \text{if (E.1}') \text{ is satisfied,} \\ C\{D_n^{(1)} + D_{n,q}^{(2)}\} & \text{if (E.2}') \text{ is satisfied.} \end{cases}$$

However, this follows from application of Corollary 2.1 to $\widetilde{X}_1, \dots, \widetilde{X}_n$ instead of $X_1, \dots, X_n$. ∎

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 [see 10, 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 26, Appendix III]; hence if the maximal eigenvalue of each $\mathrm{E}[X_i X_i']$ is bounded by a constant k_2, then every simple convex set $A \in \mathcal{A}^{\mathrm{si}}$ obeys conditions (M.2′) and (E.1′) with B_n replaced by a constant that depends only on c and k_2. Besides if the minimal eigenvalue of each $\mathrm{E}[X_i X_i']$ is bounded from below by a constant k_1, then every simple convex set $A \in \mathcal{A}^{\mathrm{si}}$ obeys condition (M.1′) with b replaced by a positive constant that depends only on k_1. Hence the conclusion of the corollary follows from application of Corollary 3.1. ∎

Proof of Corollary 3.3. Here C denotes a positive constant that depends only on b and s if condition (E.1) is satisfied, and on b, s, and q if condition (E.2) is satisfied; the value of C may change from place to place. Without loss of generality, we may assume that $B_n^2 \leq n$ since otherwise the assertions are trivial. We begin with preparing some notation. Let $R = pn^{5/2}$ and $\varepsilon = n^{-1}$, and let $\mathcal{A}_1^{\mathrm{sp}}(s)$ denote the subclass of $\mathcal{A}^{\mathrm{sp}}(s)$ consisting of every set A in $\mathcal{A}^{\mathrm{sp}}(s)$ satisfying $\max_{1 \leq j \leq p} |w_j| \leq R$ for every $w \in A$ and containing a ball with radius ε and center at, say, w_A. Let $\mathcal{A}_2^{\mathrm{sp}}(s) = \mathcal{A}^{\mathrm{sp}}(s) \setminus \mathcal{A}_1^{\mathrm{sp}}(s)$.

We divide the rest of the proof into five steps. In Steps 1-4, we verify conditions (C), (M.1′), (M.2′), (E.1′) (if (E.1) is satisfied), and condition (E.2′) (if (E.2) is satisfied) for all $A \in \mathcal{A}_1^{\mathrm{sp}}(s)$. An application of Corollary 3.1 then shows that the assertions (12) and (13) hold with $\rho_n(\mathcal{A}^{\mathrm{sp}}(s))$ replaced by $\rho_n(\mathcal{A}_1^{\mathrm{sp}}(s))$. Step 5 shows that the same assertions also hold with $\rho_n(\mathcal{A}^{\mathrm{sp}}(s))$ replaced by $\rho_n(\mathcal{A}_2^{\mathrm{sp}}(s))$. Since $\rho_n(\mathcal{A}^{\mathrm{sp}}(s)) = \rho_n(\mathcal{A}_1^{\mathrm{sp}}(s)) \vee \rho_n(\mathcal{A}_2^{\mathrm{sp}}(s))$, this will complete the proof. Step 1 relies on the following lemma, whose proof is given after the proof of this corollary.

Lemma D.1. *Let A be an s-sparsely convex set with a sparse representation $A = \cap_{q=1}^Q A_q$ for some $Q \leq p^s$. Assume that A contains the origin, that $\sup_{w \in A} \|w\| \leq R$, and that all sets A_q 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*

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

Moreover, the set A^m can be chosen to satisfy

$$\|v\|_0 \leq s \text{ for all } v \in \mathcal{V}(A^m). \tag{33}$$

Therefore, since $Q \leq p^s$, if $R \leq (pn)^{d_0}$ 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 condition (C) *for all $n \geq n_0$ with $a = 1$ and d depending only on s and d_0, and the approximating m-generated convex set A^m satisfying (33).*

Step 1. For Steps 1-4, pick any s-sparsely convex set $A \in \mathcal{A}_1^{\rm sp}(s)$ with a sparse representation $A = \cap_{q=1}^{Q} A_q$ for some $Q \leq p^s$. Here we verify condition (C) for this set A. Consider the set $B := A - w_A := \{w \in \mathbb{R}^p : w + w_A \in A\}$. The set B contains a ball with radius ε and center at the origin, satisfies the inequality $\|w\| \leq 2p^{1/2}R$ for all $w \in B$, and has a sparse representation $B = \cap_{q=1}^{Q} B_q$ where $B_q = A_q - w_A$. Clearly, each B_q satisfies $-B_q \subset \mu B_q$ with $\mu = 2p^{1/2}R/\varepsilon = 2p^{3/2}n^{7/2}$. Therefore, applying Lemma D.1 to the set B and noting that $A = B + w_A$ and $Q \leq p^s$, we see that there exists an absolute integer n_0 such that the set A satisfies condition (C) for all $n \geq n_0$ with $a = 1$ and d depending only on s, and an approximating m-generated convex set A^m such that $\|v\|_0 \leq s$ for all $v \in \mathcal{V}(A^m)$.

Step 2. Here we verify condition (M.1′). Since we have $\|v\|_0 \leq s$ for all $v \in \mathcal{V}(A^m)$, condition (M.1′) follows immediately from (M.1″).

Step 3. We shall verify condition (M.2′). For $v \in \mathcal{V}(A^m)$, let $J(v)$ be the set consisting of positions of non-zero elements of v, so that $\mathrm{Card}(J(v)) \leq s$. Using the inequality $(\sum_{j \in J(v)} |a_j|)^{2+k} \leq s^{1+k} \sum_{j \in J(v)} |a_j|^{2+k}$ for $a = (a_1, \dots, a_p)' \in \mathbb{R}^p$ (which follows from Hölder's inequality), we have for $k = 1$ or 2,

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

where $B_n' = s^3 B_n$, which leads to condition (M.2′) with B_n replaced by $s^3 B_n$.

Step 4. We shall verify condition (E.1′) when (E.1) is satisfied, and (E.2′) when (E.2) is satisfied. When (E.1) is satisfied, we have $\|X_{ij}\|_{\psi_1} \leq B_n$, so that $\|v'X_i\|_{\psi_1} \leq \sum_{j \in J(v)} \|X_{ij}\|_{\psi_1} \leq sB_n$ showing that the vectors $\widetilde{X}_i$, $i = 1, \dots, n$, satisfy (E.1′) with B_n replaced by sB_n.

When (E.2) is satisfied,

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

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

Combining Steps 1-4 and applying Corollary 3.1 shows that the assertions (12) and (13) hold with $\rho_n(\mathcal{A}^{\rm sp}(s))$ replaced by $\rho_n(\mathcal{A}_1^{\rm sp}(s))$.

Step 5. Here we show that the assertions (12) and (13) hold with $\rho_n(\mathcal{A}^{\rm sp}(s))$ replaced by $\rho_n(\mathcal{A}_2^{\rm sp}(s))$. Fix any s-sparsely convex set $A \in \mathcal{A}_2^{\rm sp}(s)$ with a sparse representation $A = \cap_{q=1}^{Q} A_q$ for some $Q \leq p^s$. Let $A^R := \{w \in$

$\mathbb{R}^p : \max_{1\le j\le p}|w_j| > R\}$. Then $A = \overline{A} \cup (A \cap A^R)$ for some s-sparsely convex set $\overline{A} \subset \mathbb{R}^p$ such that $\max_{1\le j\le p}|w_j| \le R$ for all $w \in \overline{A}$.

Now observe that by Markov's inequality,

$$\mathrm{P}\left(\max_{i,j}|X_{ij}| > pn^2\right) \le \frac{\mathrm{E}[\max_{i,j}|X_{ij}|]}{pn^2} \le \frac{\mathrm{E}[\sum_{i,j}|X_{ij}|]}{pn^2}$$
$$\le \max_{i,j}\mathrm{E}[|X_{ij}|]/n \le CB_n/n \le C/n^{1/2},$$

where $\max_{i,j}$ stands for $\max_{1\le i\le n}\max_{1\le j\le p}$. Hence

$$\mathrm{P}\left(S_n^X \in A^R\right) \le C/n^{1/2}.$$

It is easy to verify that the same inequality also holds with S_n^X replaced by S_n^Y, and we have

$$|\mathrm{P}(S_n^X \in A) - \mathrm{P}(S_n^Y \in A)| \le |\mathrm{P}(S_n^X \in \overline{A}) - \mathrm{P}(S_n^Y \in \overline{A})| + C/n^{1/2}.$$

Therefore it suffices to only consider the case where $A \in \mathcal{A}_2^{\mathrm{sp}}(s)$ is such that $\max_{1\le j\le p}|w_j| \le R$ for all $w \in A$.

Next, we consider the two cases separately. First, suppose that at least one A_q does not contain a ball with radius ε. Then the set $\cap_{v\in\mathbb{S}^{p-1}}\{w \in \mathbb{R}^p : w'v \le S_{A_q}(v) - \varepsilon\}$ is empty, and so under condition (M.2), Lemma A.2 implies that $\mathrm{P}(S_n^Y \in A_q) \le C\varepsilon = C/n$ (since the Hilbert-Schmidt norm is equal to the square-root of the sum of squares of the eigenvalues of the matrix, under our condition (M.1″), the constant C in the bound $C\varepsilon$ above depends only on b and s). In addition, under conditions (M.1″) and (M.2), the Berry-Esseen theorem [see 21, Theorem 1.3] implies that

$$\left|\mathrm{P}(S_n^X \in A_q) - \mathrm{P}(S_n^Y \in A_q)\right| \le CB_n/n^{1/2}.$$

Since $A \subset A_q$, both $\mathrm{P}(S_n^X \in A)$ and $\mathrm{P}(S_n^Y \in A)$ are bounded from above by the quantities on the right hand sides of (12) and (13) depending on whether (E.1) or (E.2) is satisfied, and so is their difference. This completes the proof in this case.

Second, suppose that each A_q contains a ball with radius ε (possibly depending on q). Then applying Lemma D.1 to each A_q separately shows that for $n \ge n_0$ and $m \le (pn)^d$ with d depending only on s, we can construct an m-generated convex sets A_q^m such that

$$A_q^m \subset A_q \subset A_q^{m,1/n}$$

and $\|v\|_0 \le s$ for all $v \in \mathcal{V}(A_q^m)$. The set $A^0 = \cap_{q=1}^Q A_q^{m,1/n}$ trivially satisfies condition (C) with $a = 0$ and d depending only on s. In addition, it follows from the same arguments as those used in Steps 2-4 that the set A^0 satisfies conditions (M.1′), (M.2′), (E.1′) (if (E.1) is satisfied), and (E.2′) (if (E.2) is satisfied). Therefore, by applying Corollary 3.1, we conclude that $|\mathrm{P}(S_n^X \in A^0) - \mathrm{P}(S_n^Y \in A^0)|$ is bounded from above by the quantities on the right hand sides of (10) and (11) depending on whether (E.1) or (E.2) is satisfied. Also, observe that $A \subset A^0$ and that $\cap_{q=1}^Q A_q^{m,-\varepsilon}$ is empty because $\cap_{q=1}^Q A_q^m \subset A$

and A contains no balls with radius ε. This implies that $\mathrm{P}(S_n^Y \in A^0) \leq C(\log p)^{1/2}/n$ by Lemma A.1 and condition (M.1′). Since $A \subset A^0$, both $\mathrm{P}(S_n^X \in A)$ and $\mathrm{P}(S_n^Y \in A)$ are bounded from above by the quantities on the right hand sides of (12) and (13) depending on whether (E.1) or (E.2) is satisfied, and so is their difference. This completes the proof in this case. ■

Here we prove Lemma D.1 used in the proof of Corollary 3.3.

Proof of Lemma D.1. For convex sets P_1 and P_2 containing the origin and such that $P_1 \subset P_2$, define

$$d_{BM}(P_1, P_2) := \inf\{\epsilon > 0 : P_2 \subset (1+\epsilon)P_1\}.$$

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 $P_3 \subset P_4$,

$$d_{BM}(P_1 \cap P_3, P_2 \cap P_4) \leq d_{BM}(P_1 \cap P_2) \vee d_{BM}(P_3 \cap P_4). \tag{34}$$

Let $A = \cap_{q=1}^Q A_q$ be a sparse 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 $s_q \leq s$ elements of its argument $w = (w_1, \dots, w_p)$. Since A contains the origin, A_q contains the origin as well. Therefore, applying Corollary 1.5 in [4] as if A_q was a set in $\mathbb{R}^{s_q}$ shows that one can construct a polytope $P_q \subset \mathbb{R}^p$ with at most $(\gamma((\mu+1)/\epsilon)^{1/2}\log(1/\epsilon))^{s_q}$ vertices such that

$$P_q \subset A_q \subset (1+\epsilon)P_q$$

and such that for all $v \in \mathcal{V}(P_q)$, non-zero elements of v correspond to some of the main components of A_q. Since we need at most s_q vertices to form a face of the polytope P_q, the polytope P_q has

$$m_q \leq \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s_q^2} \leq \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s^2} \tag{35}$$

faces. Now observe that P_q is an m_q-generated convex set. Thus, we have constructed 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 elements. Hence

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

Next, it follows from (34) 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 first claim of the lemma now follows from (35). The second claim (33) holds by construction of A^m, and the final claim is trivial. ■

Appendix E. Proofs for Section 4

E.1. **Maximal inequalities.** Here we collect some useful maximal inequalities that will be used in the proofs for Section 4.

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

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

where K is a universal constant.

Proof. See Lemma 8 in [18]. ■

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

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

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

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

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 [20]. ■

Lemma E.3. *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 [18]. ■

Lemma E.4. *Assume the setting of Lemma E.3. (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 E.4 relies on the following lemma, which follows from Theorem 10 in [25].

Lemma E.5. *Assume the setting of Lemma E.3. 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 E.5. By homogeneity, we may assume that $B = 1$. Then by Theorem 10 in [25], 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 hand 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 E.4. 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 24, Proposition 6.8], we have

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

Hence

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

By Lemma E.5, 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 [24] (note that a version of their theorem applies to nonnegative random vectors) and the fact that $\mathrm{E}[W_2] \le 8\mathrm{E}[M]$,

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

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

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

Case (ii). By Theorem 6.20 in [24] (note that a version of their theorem applies to nonnegative random vectors) and the fact that $\mathrm{E}[W_2] \le 8\mathrm{E}[M]$,

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

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

Proof of Theorem 4.1. In this proof, C is a positive constant that depends only on a, b, and d but its value may change at each appearance. Fix any $A \in \mathcal{A}^{\mathrm{si}}$. Let A^m be an approximating m-generated convex set as in (C). By assumption, $A^m \subset A \subset A^{m,\epsilon}$. Let

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

As in the proof of Corollary 3.1, we have

$$\begin{aligned}&|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)|\\ &\quad \le C\epsilon\log^{1/2}(pn) + \overline{\rho} \le Cn^{-1}\log^{1/2}(pn) + \overline{\rho},\end{aligned}$$

so that the problem reduces to proving that under (M.1), the inequality

$$\rho_n^{MB}(\mathcal{A}^{\mathrm{re}}) \le C\bar{\Delta}_n^{1/3}\log^{2/3} p \tag{36}$$

holds on the event $\Delta_{n,r} \le \bar{\Delta}_n$, where $\Delta_{n,r} := \max_{1\le j,k\le p}|\Sigma_{jk}^{eX} - \Sigma_{jk}^Y|$ with Σ_{jk}^{eX} and Σ_{jk}^Y denoting the (j,k)-th elements Σ^{eX} and Σ^Y, respectively.

To this end, we first show that

$$\varrho_n^{MB} := \sup_{y\in\mathbb{R}^p}|\mathrm{P}(S_n^{eX} \le y \mid X_1^n) - \mathrm{P}(S_n^Y \le y)| \le C\Delta_{n,r}^{1/3}\log^{2/3} p. \tag{37}$$

To show (37), fix any $y = (y_1, \dots, y_p)' \in \mathbb{R}^p$. As in the proof of Lemma 5.1, for $\beta > 0$, define

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

Note that conditional on X_1^n, S_n^{eX} is a centered Gaussian random vector with covariance matrix Σ_n^{eX}. Then a small modification of the proof of Theorem 1 in [18] implies that for every $g \in C^2(\mathbb{R})$ with $\|g'\|_\infty \vee \|g''\|_\infty < \infty$, we have

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

Hence, as in Step 2 of the proof of Lemma 5.1, we obtain with $\phi = \beta / \log p$ that

$$\begin{aligned} &|\mathrm{P}(S_n^{eX} \leq s - \phi^{-1} \mid X_1^n) - \mathrm{P}(S_n^Y \leq s - \phi^{-1})| \\ &\leq C \left\{ \phi^{-1} (\log p)^{1/2} + (\phi^2 + \beta\phi)\Delta_{n,r} \right\}. \end{aligned}$$

Substituting $\beta = \phi \log p$, optimizing the resulting expression with respect to ϕ, and noting that $y \in \mathbb{R}^p$ is arbitrary lead to (37). Finally (36) follows from the fact that the inequality $\varrho_n^{MB} \leq C \bar{\Delta}_n^{1/3} \log^{2/3} p$ holds on the event $\Delta_{n,r} \leq \bar{\Delta}_n$, and applying the same argument as that used in the proof of Corollary 5.1. ■

Proof of Corollary 4.1. In this proof, c and C are positive constants that depend only on a, b, d, and s under (E.1), and on a, b, d, s, and q under (E.2); their values may vary from place to place. For brevity of notation, we implicitly assume here 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

$$B_n^2 (\log^5 (pn)) \log^2 (1/\alpha) \leq n \tag{38}$$

since otherwise the assertions are trivial.

We shall apply Theorem 4.1 to prove the corollary. Since $n^{-1/2} \log^{1/2}(pn) \leq C D_n^{(1)}(\alpha)$, it suffices to construct an appropriate $\bar{\Delta}_n$ such that $\mathrm{P}(\Delta_n > \bar{\Delta}_n) \leq \alpha$ and to bound $\bar{\Delta}_n^{1/3} \log^{2/3}(pn)$.

We begin with observing that under condition (C$'$), $\Delta_n \leq C \Delta_{n,r}$ where $\Delta_{n,r} = \max_{1 \leq j,k \leq p} |\Sigma_{jk}^{eX} - \Sigma_{jk}^Y|$. As $\mathrm{E}[X_i X_i'] = \mathrm{E}[Y_i Y_i']$ for all i, we have

$$\Sigma^{eX} - \Sigma^Y = n^{-1} \sum_{i=1}^n (X_i X_i' - \mathrm{E}[X_i X_i']) - \widehat{\mu}_n^X (\widehat{\mu}_n^X)',$$

by which we have $\Delta_{n,r} \leq \Delta_{n,r}^{(1)} + \{\Delta_{n,r}^{(2)}\}^2$, where

$$\Delta_{n,r}^{(1)} := \max_{1 \leq j,k \leq p} \left| n^{-1} \sum_{i=1}^n (X_{ij} X_{ik} - \mathrm{E}[X_{ij} X_{ik}]) \right|, \ \Delta_{n,r}^{(2)} := \max_{1 \leq j \leq p} \left| \widehat{\mu}_{nj}^X \right|.$$

The desired assertions then follow from the bounds on $\Delta^{(1)}_{n,r}$ and $\Delta^{(2)}_{n,r}$ derived separately for (E.1) and (E.2) cases below.

Case (E.1). Observe that by Hölder's inequality and (M.2),

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

In addition, by (E.1),

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

so that for $M_n := \max_{i,j,k} |X_{ij}X_{ik} - \mathrm{E}[X_{ij}X_{ik}]|$, we have

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

which also implies that $(\mathrm{E}[M_n^2])^{1/2} \leq CB_n^2 \log^2(pn)$. Hence by Lemma E.1, we have

$$\begin{aligned} \mathrm{E}[\Delta_{n,1}] &\leq Cn^{-1}\{\sqrt{\sigma_n^2 \log p} + \sqrt{\mathrm{E}[M_n^2]} \log p\} \\ &\leq C\{(n^{-1}B_n^2 \log p)^{1/2} + n^{-1}B_n^2 \log^3(pn)\} \leq C\{n^{-1}B_n^2 \log(pn)\}^{1/2}, \end{aligned}$$

where the last inequality follows from (38). Applying Lemma E.2 (i) with $\beta = 1/2$ and $\eta = 1$, we conclude that for every $t > 0$,

$$\begin{aligned} &\mathrm{P}\left\{\Delta^{(1)}_{n,r} > C\{n^{-1}B_n^2 \log(pn)\}^{1/2} + t\right\} \\ &\quad \leq \exp\{-nt^2/(3B_n^2)\} + 3\exp\{-c\sqrt{nt}/(B_n \log(pn))\}. \end{aligned}$$

Choosing $t = C\{n^{-1}B_n^2(\log(pn))\log^2(1/\alpha)\}^{1/2}$ for sufficiently large $C > 0$, the right hand side is bounded by

$$\alpha/4 + 3\exp\{-cC^{1/2}n^{1/4}(\log^{1/2}(1/\alpha))/(B_n^{1/2}\log^{3/4}(pn))\} \leq \alpha/2,$$

where the last inequality follows from (38). Therefore

$$\mathrm{P}\{\{\Delta^{(1)}_{n,r} \log^2 p\}^{1/3} > CD_n^{(1)}(\alpha)\} \leq \alpha/2.$$

It is routine to verify that the same inequality holds with $\Delta^{(1)}_{n,r}$ replaced by $\{\Delta^{(2)}_{n,r}\}^2$. This leads to the conclusion of the corollary under (E.1).

Case (E.2). Define σ_n^2 and M_n by the same expressions as those in the previous case; then $\sigma_n^2 \leq nB_n^2$. For M_n, we have

$$\begin{aligned} \mathrm{E}[M_n^{q/2}] &\leq C\{\mathrm{E}[\max_{i,j,k} |X_{ij}X_{ik}|^{q/2}] + \max_{i,j,k}(\mathrm{E}[|X_{ij}X_{ik}|])^{q/2}\} \\ &\leq C\{\mathrm{E}[\max_{i,j,k} |X_{ij}X_{ik}|^{q/2}]\} = C\mathrm{E}[\max_{i,j} |X_{ij}|^q] \leq CnB_n^q, \end{aligned}$$

which also implies that $(\mathrm{E}[M_n^2])^{1/2} \le Cn^{2/q}B_n^2$. Hence by Lemma E.1, we have

$$\begin{aligned}\mathrm{E}[\Delta_{n,r}^{(1)}] &\le Cn^{-1}\{\sqrt{\sigma_n^2 \log p} + \sqrt{\mathrm{E}[M_n^2]} \log p\} \\ &\le C\{(n^{-1}B_n^2 \log p)^{1/2} + n^{-1+2/q}B_n^2 \log p\}.\end{aligned}$$

Applying Lemma E.2 (ii) with $s = q/2$ and $\eta = 1$, we have for every $t > 0$,

$$\begin{aligned}&\mathrm{P}\left\{\Delta_{n,r}^{(1)} > C\{(n^{-1}B_n^2 \log p)^{1/2} + n^{-1+2/q}B_n^2 \log p\} + t\right\} \\ &\quad \le \exp\{-nt^2/(3B_n^2)\} + ct^{-q/2}n^{1-q/2}B_n^q.\end{aligned}$$

Choosing

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

for sufficiently large $C > 0$, we conclude that

$$\mathrm{P}\{\{\Delta_{n,r}^{(1)} \log^2 p\}^{1/3} > C\{D_n^{(1)}(\alpha) + D_{n,q}^{(2)}(\alpha)\}\} \le \alpha/2.$$

It is routine to verify that the same inequality holds with $\Delta_{n,r}^{(1)}$ replaced by $\{\Delta_{n,r}^{(2)}\}^2$. This leads to the conclusion of the corollary under (E.2). ■

Proof of Corollary 4.2. Here C is understood to be a positive constant that depends only on a, d, k_1 and k_2; the value of C may change from place to place. To prove this corollary, we apply Theorem 4.1, to which end we have to verify condition (M.1′) and derive a suitable bound on Δ_n. Condition (M.1′) follows from the fact that the minimum eigenvalue of $\mathrm{E}[X_i X_i']$ is bounded from below by k_1. By log-concavity of the distributions of X_i, we have $\|v'X_i\|_{\psi_1} \le C(\mathrm{E}[(v'X_i)^2])^{1/2} \le C$ for all $v \in \mathbb{R}^p$ with $\|v\| = 1$ (see the proof of Corollary 3.2). For each $1 \le i \le n$, let $\check{X}_i$ be a random vector whose elements are given by $v'X_i, v \in \cup_{A \in \mathcal{A}^{\mathrm{si}}} \mathcal{V}(A^m(A))$; for each $1 \le i \le n$, the dimension of $\check{X}_i$, denoted by $\check{p}$, is at most $(pn)^d$, and$\|\check{X}_{ij}\|_{\psi_1} \le C$ for all $1 \le j \le \check{p}$. Then Δ_n coincides with $\Delta_{n,r}$ with X_i replaced by $\check{X}_i$, that is,

$$\Delta_n = \max_{1 \le j,k \le \check{p}} \left| n^{-1} \sum_{i=1}^n (\check{X}_{ij}\check{X}_{ik} - \mathrm{E}[\check{X}_{ij}\check{X}_{ik}]) - \mathbb{E}_n[\check{X}_{ij}]\mathbb{E}_n[\check{X}_{ik}] \right|.$$

Noting that $\log \check{p} \le d \log(pn)$, by the same argument as that used in the proof of Corollary 4.1 case (E.1), we can find a constant $\bar{\Delta}_n$ such that $\mathrm{P}(\Delta_n > \bar{\Delta}_n) \le \alpha$ and

$$(\bar{\Delta}_n \log^2 p)^{1/3} \le C\{n^{-1}(\log^5(pn)) \log^2(1/\alpha)\}^{1/6}.$$

Here without loss of generality we assume that $(\log^5(pn)) \log^2(1/\alpha) \le n$. The desired assertion then follows. ■

Proof of Corollary 4.3. In this proof, let C be a positive constant depending only on b, s, q (C depends on q only in the case where (E.2) is satisfied); the value of C may change from place to place. Moreover, without loss of generality, we will assume that

$$B_n^2(\log^5(pn))\log^2(1/\alpha) \le n$$

since otherwise the assertions are trivial.

Let $\Delta_{n,r} := \max_{1\le j,k\le p}|\Sigma^{eX}_{jk} - \Sigma^{Y}_{jk}|$, and

$$\bar{\Delta}_n = \begin{cases} \left(\frac{B_n^2(\log(pn))\log^2(1/\alpha)}{n}\right)^{1/2} & \text{if (E.1) is satisfied} \\ \left(\frac{B_n^2(\log(pn))\log^2(1/\alpha)}{n}\right)^{1/2} + \frac{B_n^2\log p}{\alpha^{2/q}n^{1-q/2}} & \text{if (E.2) is satisfied.} \end{cases}$$

Then by the proof of Corollary 4.1, in either case where (E.1) or (E.2) is satisfied, there exists a positive constant C_1 depending only on b, s, q (C_1 depends on q only in the case where (E.2) is satisfied) such that

$$\mathrm{P}(\Delta_{n,r} > C_1\bar{\Delta}_n) \le \alpha/2.$$

We may further assume that $C_1 s\bar{\Delta}_n \le b/2$, since otherwise the assertions are trivial.

Let $R = pn^{5/2}$ and $\varepsilon = n^{-1}$, and define the subclasses $\mathcal{A}_1^{\mathrm{sp}}(s)$ and $\mathcal{A}_2^{\mathrm{sp}}(s)$ as in the proof of Corollary 3.3. Moreover, as in the proof of Corollary 3.3, for any $A \in \mathcal{A}^{\mathrm{sp}}(s)$, we can verify conditions (C), (M.1′), (M.2′), (E.1′) (if (E.1) is satisfied), and condition (E.2′) (if (E.2) is satisfied). Therefore, the bounds (14) and (15) with $\rho_n^{MB}(\mathcal{A}^{\mathrm{sp}}(s))$ replaced by $\rho_n^{MB}(\mathcal{A}_1^{\mathrm{sp}}(s))$ follow from Corollary 4.1. Since $\rho_n^{MB}(\mathcal{A}^{\mathrm{sp}}(s)) = \rho_n^{MB}(\mathcal{A}_1^{\mathrm{sp}}(s)) \vee \rho_n^{MB}(\mathcal{A}_2^{\mathrm{sp}}(s))$, it remains to bound $\rho_n^{MB}(\mathcal{A}_2^{\mathrm{sp}}(s))$.

As in the proof of Corollary 3.3, fix any s-sparsely convex set $A \in \mathcal{A}_2^{sp}(s)$ with sparse representation $A = \cap_{q=1}^{Q} A_q$ with $Q \le p^s$, and let $A^R = \{w \in A : \max_{1\le j\le p}|w_j| > R\}$. Then $A = \overline{A} \cup (A \cap A^R)$ for some s-sparsely convex set $\overline{A}$ with $\max_{1\le j\le p}|w_j| \le R$ for all $w \in \overline{A}$. It is routine to verify that $\mathrm{P}(S_n^Y \in A^R) \le C/n^{1/2}$. Moreover, conditional on X_1^n, S_{nj}^{eX} is Gaussian with mean zero and variance $\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj}^X)^2] = \Sigma_{jj}^{eX}$, so that

$$\begin{aligned} \mathrm{P}(S_n^{eX} \in A^R \mid X_1^n) &\le \mathrm{P}(\max_{1\le j\le p}|S_{nj}^{eX}| > R \mid X_1^n) \\ &\le \frac{\mathrm{E}[\max_{1\le j\le p}|S_{nj}^{eX}| \mid X_1^n]}{R} \le \frac{C(\log p)^{1/2}\max_{1\le j\le p}(\Sigma_{jj}^{eX})^{1/2}}{R}, \end{aligned}$$

which is bounded by $C/n^{1/2}$ on the event $\Delta_{n,r} \le C_1\bar{\Delta}_n$. Hence on the event $\Delta_{n,r} \le C_1\bar{\Delta}_n$,

$$\begin{aligned} &|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| \\ &\le |\mathrm{P}(S_n^{eX} \in \overline{A} \mid X_1^n) - \mathrm{P}(S_n^Y \in \overline{A})| + C/n^{1/2}, \end{aligned}$$

so that it suffices to only consider the case where $A \in \mathcal{A}_2^{sp}(s)$ is such that $\max_{1\le j\le p}|w_j| \le R$ for all $w \in A$.

As in the proof of Corollary 3.3, we separately consider two cases. First, suppose that at least one of A_q does not contain a ball of radius ε; then $\cap_{v \in \mathbb{S}^{p-1}} \{w \in \mathbb{R}^p : w'v \le \mathcal{S}_{A_q}(v) - \varepsilon\}$ is empty, so that by condition (M.1$''$) and Lemma A.2, $\mathrm{P}(S_n^Y \in A_q) \le C\varepsilon$. Moreover, since S_n^{eX} is Gaussian conditional on X_1^n, by condition (M.1$''$) and Lemma A.2, we have, on the event $\Delta_{n,r} \le C_1 \bar{\Delta}_n$, $\mathrm{P}(S_n^{eX} \in A_q \mid X_1^n) \le C\varepsilon$. Since $A \subset A_q$, we conclude that on the event $\Delta_{n,r} \le C_1 \bar{\Delta}_n$, $|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| \le C\varepsilon = C/n$.

Second, suppose that each A_q contains a ball with radius ε. Then by applying Lemma D.1 to each A_q, for $n \ge n_0$ and $m \le (pn)^d$ with d depending only on s, we can construct an m-generated convex set A_q^m such that $A_q^m \subset A_q \subset A_q^{m,1/n}$ with $\|v\|_0 \le s$ for all $v \in \mathcal{V}(A_q^m)$. Let $A_0 = \cap_{q=1}^Q A_q^{m,1/m}$; then $A \subset A^0$ and $\cap_{q=1}^Q A_q^{m,-\varepsilon}$ is empty. By the latter fact, together with condition (M.1$''$) and Lemma A.1, we have $\mathrm{P}(S_n^Y \in A^0) \le C(\log p)^{1/2}/n$. Moreover, since S_n^{eX} is Gaussian conditional on X_1^n, by condition (M.1$''$) and Lemma A.1, the inequality $\mathrm{P}(S_n^{eX} \in A^0 \mid X_1^n) \le C(\log p)^{1/2}/n$ holds on the event $\Delta_{n,r} \le C_1 \bar{\Delta}_n$. Since $A \subset A^0$, we conclude that on the event $\Delta_{n,r} \le C_1 \bar{\Delta}_n$, $|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| \le C(\log p)^{1/2}/n$. The final conclusion follows from the fact that $\mathrm{P}(\Delta_{n,r} > C_1 \bar{\Delta}_n) \le \alpha/2$. ■

Proof of Theorem 4.2. By the triangle inequality, $\rho_n^{EB} \le \rho_n^{MB} + \varrho_n^{EB}$, where

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

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 matrix Σ_n^{eX}. In addition, conditional on X_1^n, $S_n^{eX} \stackrel{d}{=} \sum_{i=1}^n Y_i^*/\sqrt{n}$, where $Y_1^*, \dots, Y_n^*$ are i.i.d. centered Gaussian random vectors with the same covariance matrix Σ_n^{eX}. Hence the conclusion of the theorem follows from 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 $\{\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj})^2] \ge b \text{ for all } 1 \le j \le p\} \cap \{\widehat{L}_n \le \overline{L}_n\} \cap \{\widehat{M}_n(\phi_n) \le \overline{M}_n\}$. ■

Proof of Corollary 4.4. 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

$$B_n^2 \log^7(pn) \le cn, \tag{39}$$

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

Moreover, by the same argument as that used in the proof of Corollary 4.1, the problem reduces to the case of rectangles $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$; that is, it suffices to prove the bounds (16) and (17) with $\rho_n^{EB}(\mathcal{A}^{\mathrm{si}})$ replaced by $\rho_n^{EB}(\mathcal{A}^{\mathrm{re}})$ and condition (M.1$'$) replaced by (M.1). For the latter problem, we will apply Theorem 4.2.

Case (E.1). With (39) 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 $|a-b|^3 \leq 4(|a|^3 + |b|^3)$ together with Jensen's inequality, we have

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

By Lemma E.3,

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

Note that $\| |X_{ij}|^3 \|_{\psi_{1/3}} \leq \|X_{ij}\|_{\psi_1}^3 \leq B_n^3$, so that applying Lemma E.4 (i) with $\beta = 1/3$, we have for every $t > 0$,

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

Taking $t = \log(18/\alpha) \leq C \log(pn)$, we conclude that, with $\overline{L}_n = CB_n$ (recall (39)), $\mathrm{P}(\widehat{L}_n > \overline{L}_n) \leq \alpha/6$.

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

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

so that

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

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

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

Observe that

$$B_n \phi_n \log p \leq C n^{-1/6} B_n^{2/3} \log^{1/3} p,$$

so that using (39), we conclude that $\mathrm{P}\{\widehat{M}_{n,X}(\phi_n) > 0\} \leq \alpha/6$. For $\widehat{M}_{n,Y}(\phi_n)$, since with probability larger than $1-\alpha/6$, $\mathbb{E}_n[(X_{ij} - \widehat{\mu}_{nj}^X)^2] \leq CB_n$ for all $j = 1, \dots, p$, on that event, conditional on $X_1, \dots, X_n$, $\|S_{nj}^{eX}\|_{\psi_2} \leq CB_n^{1/2}$ for all $j = 1, \dots, p$. Hence, using the same argument used in bounding $\widehat{M}_{n,X}(\phi_n)$, we conclude that

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

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

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

The final conclusion follows from Corollary 4.1.

Case (E.2). In this case, in addition to (39), we may assume that

$$\frac{B_n^2 \log^3(pn)}{\alpha^{2/q} n^{1-2/q}} \leq c \leq 1, \tag{40}$$

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 E.3,

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

Hence by applying Lemma E.4 (ii) with $s = q/3$, we have for every $t > 0$,

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

Solving $Ct^{-q/3} n B_n^q = \alpha/6$, 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 hand side is nondecreasing in ϕ_n, and

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

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

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

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

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

and because of (40), the second term inside the bracket on the right hand side is at most

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

■

Proof of Corollary 4.5. The proof is analogous to that of Corollary 4.2 ■

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] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411-420

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

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

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

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

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

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

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

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

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

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

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

[15] 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.

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

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

[18] 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.*
[19] Dudley, R.M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press.
[20] Einmahl, U. and Li, D. (2008). Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.* **360** 6677-6693.
[21] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724-739.
[22] Goldstein, L., and Rinott, Y. (1996) Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1-17.
[23] 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.*
[24] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer.
[25] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863-884.
[26] Milman, V.D. and Shechtman, G. (1986). *Asymptotic Theory of Finite Dimensional Normed Spaces.* Lecture Notes in Mathematics 1200 Springer.
[27] 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.
[28] 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.
[29] Panchenko, D. (2013). *The Sherrington-Kirkpatrick Model.* Springer.
[30] Praestgaard, J. and Wellner, J.A. (1993). Exchangeably weighted bootstraps of the general empirical processes. *Ann. Probab.* **21** 2053-2086.
[31] 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.
[32] Röllin, A. (2011). Stein's method in high dimensions with applications, *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 529-549.
[33] Sazonov, V. (1968). On the multi-dimensional central limit theorem. *Sanhya Ser. A* **30** 181-204.
[34] Sazonov, V. (1981). *Normal Approximations: Some Recent Advances.* Lecture Notes in Mathematics Volume 879, Springer.
[35] Senatov, V. (1980). Several estimates of the rate of convergence in the multidimensional CLT. *Dokl. Acad. Nauk SSSR* **254** 809-812.
[36] Sweeting, T. (1977). Speed of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28-41.

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

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

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

[40] 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