The Annals of Probability
2017, Vol. 45, No. 4, 2309–2352
DOI: 10.1214/16-AOP1113
© Institute of Mathematical Statistics, 2017

CENTRAL LIMIT THEOREMS AND BOOTSTRAP IN HIGH DIMENSIONS

BY VICTOR CHERNOZHUKOV[1], DENIS CHETVERIKOV AND KENGO KATO[2]

Massachusetts Institute of Technology, University of California, Los Angeles and University of Tokyo

This paper derives central limit and bootstrap theorems for probabilities that sums of centered high-dimensional random vectors hit hyperrectangles and sparsely convex sets. Specifically, we derive Gaussian and bootstrap approximations for 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 hyperrectangle, or more generally, a sparsely convex set, and show that the approximation error converges to zero even if $p = p_n \to \infty$ as $n \to \infty$ 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 hyperrectangles, or more generally, sparsely convex sets, and does not require any restriction on the correlation structure among coordinates of X_i. Sparsely convex sets are sets that can be represented as intersections of many convex sets whose indicator functions depend only on a small subset of their arguments, with hyperrectangles being a special case.

1. Introduction.

Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ where $p \geq 3$ may be large or even much larger than n. Denote by X_{ij} the jth 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:

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

Received April 2015; revised March 2016.

[1]Supported by a National Science Foundation grant.

[2]Supported by the Grant-in-Aid for Young Scientists (B) (25780152), the Japan Society for the Promotion of Science.

MSC2010 subject classifications. 60F05, 62E17.

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

We are interested in bounding the quantity

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

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 the bounds, has been studied by a number of authors; see, for example, [5, 7, 9, 24, 29, 35–37] and [40]; we refer to [18] for an exhaustive literature review. Typically, we are interested in how fast $p = p_n \to \infty$ is allowed to grow while guaranteeing $\rho_n(\mathcal{A}) \to 0$. In particular, Bentkus [5] established one of the sharpest results in this direction which states that when $X_1, \dots, X_n$ are i.i.d. with $\mathrm{E}[X_i X_i'] = I$ (I denotes the $p \times p$ identity matrix),

$$\rho_n(\mathcal{A}) \le C_p(\mathcal{A}) \frac{\mathrm{E}[\|X_1\|^3]}{\sqrt{n}}, \tag{2}$$

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 $C_p(\mathcal{A}) \le 400p^{1/4}$ when $\mathcal{A}$ is the class of all Borel measurable 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 Jensen's inequality, when $\mathrm{E}[X_1 X_1'] = I$, $\mathrm{E}[\|X_1\|^3] \ge (\mathrm{E}[\|X_1\|^2])^{3/2} = p^{3/2}$, 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 Borel measurable 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 Borel measurable convex sets, it was proved by [29] that $\rho_n(\mathcal{A}) \ge c\mathrm{E}[\|X_1\|^3]/\sqrt{n}$ 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 hyperrectangles, or more generally for $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$ being a class of simple convex sets, and show that these bounds lead to results of type (3). We call any convex set a simple convex set if it can be well approximated by a convex polytope whose number of facets is (potentially very large but) not too large; see Section 3 for details. 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 to statistics. In particular, the results for hyperrectangles and sparsely convex sets are of importance because they allow us to approximate the distributions of various key statistics that arise in inference for high-dimensional models. For example, the probability that a collection of Kolmogorov–Smirnov type statistics falls below a collection of thresholds

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

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 (nonintersecting) subsets of $\{1,\dots,p\}$, $\{t_k\}$ are thresholds in the interval $(-\infty,\infty)$, $\kappa \ge 1$ is an integer, and $A \in \mathcal{A}^{\mathrm{re}}$ is a hyperrectangle 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

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

which 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 $(0,\infty)$, $\kappa \ge 1$ 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 [17] 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, as the main ingredient in the new proof, we employ an argument inspired by Bolthausen [10]. Our paper builds upon our previous work [17], which in turn builds on a number of works; see [13–15, 21, 23, 31, 33, 34, 38, 39] and [41] (see also [18] 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 hyperrectangles 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 hyperrectangles 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 $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$ being a class of simple convex sets and for $\mathcal{A} = \mathcal{A}^{\mathrm{sp}}(s)$ being the class of all s-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 $A \in \mathcal{A}^{\mathrm{re}}$, $\mathcal{A}^{\mathrm{si}}(a,d)$, or $\mathcal{A}^{\mathrm{sp}}(s)$ using the data $X_1,\dots,X_n$. In Section 5, we state an important technical lemma, which constitutes the main part of the derivation of our high-dimensional CLT. Finally, we provide all the 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 $\|v\|_0 := \sum_{j=1}^p 1\{v_j \ne 0\}$ 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

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

For $\alpha \in [1, \infty)$, $\|\cdot\|_{\psi_\alpha}$ is an Orlicz norm, while for $\alpha \in (0, 1)$, $\|\cdot\|_{\psi_\alpha}$ is not a norm but a quasi-norm, that is, there exists a constant K_α depending only on α such that $\|\xi_1 + \xi_2\|_{\psi_\alpha} \le K_\alpha(\|\xi_1\|_{\psi_\alpha} + \|\xi_2\|_{\psi_\alpha})$. Throughout the paper, we assume that $n \ge 4$ and $p \ge 3$.

2. High-dimensional CLT for hyperrectangles.

This section presents a high-dimensional CLT for hyperrectangles. We begin with presenting an abstract theorem (Theorem 2.1); the bound in Theorem 2.1 is general but depends on the tail properties of the distributions of the coordinates of X_i in a nontrivial way. Next, we apply this theorem under simple moment conditions and derive more explicit bounds (Proposition 2.1).

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

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

for some $-\infty \le a_j \le b_j \le \infty$, $j = 1, \dots, p$. We will derive a bound on $\rho_n(\mathcal{A}^{\mathrm{re}})$, 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

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

and for $\phi \ge 1$, define

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

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 hyperrectangles). *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 on b such that for every constant $\overline{L}_n \geq L_n$, we have*

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

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 [5] 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

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

for some universal constant K; see Proposition 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 $\sqrt{p} = o(n^{1/7})$ to make $\rho_n(\mathcal{A}) \to 0$ when $\mathcal{A}$ is the class of all Borel measurable 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, replacing $\sqrt{p}$ by $\log p$.

(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 [6]), 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 [19] suggest that dependence on B_n is also optimal. Hence, we conjecture that up to a universal constant,

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

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 [17] 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 [17] 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 [17] to the class of hyperrectangles $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$ 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 $i = 1, \dots, n$.

In addition, denote

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

An application of Theorem 2.1 under these conditions leads to the following proposition.

PROPOSITION 2.1 (High-dimensional CLT for hyperrectangles). *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 (Proposition 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 classes of sparsely convex sets. These classes in turn may be of interest in statistics where sparse models and techniques have been of canonical importance in the past years.

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

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

where $\mathbb{S}^{p-1} := \{v \in \mathbb{R}^p : \|v\| = 1\}$; in particular, $A = \bigcap_{v \in \mathbb{S}^{p-1}} \{w \in \mathbb{R}^p : w'v \leq \mathcal{S}_A(v)\}$. We say that the set A is *m-generated* if it is generated by the intersection of m half-spaces (that is, A is a convex polytope with at most m facets). 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,

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

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

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

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 $\mathcal{A}^{\mathrm{si}}(a,d)$ be the class of all Borel sets A in $\mathbb{R}^p$ that satisfy the following condition:

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

We refer to sets A that satisfy condition (C) as *simple convex sets*. Note that any hyperrectangle $A \in \mathcal{A}^{\mathrm{re}}$ satisfies condition (C) with $a = 0$ and $d = 1$ (recall that $n \geq 4$), and so belongs to the class $\mathcal{A}^{\mathrm{si}}(0,1)$. For $A \in \mathcal{A}^{\mathrm{si}}(a,d)$, let $A^m(A)$ denote the corresponding set A^m that appears in condition (C).

We will consider subclasses $\mathcal{A}$ of the class $\mathcal{A}^{\mathrm{si}}(a,d)$ consisting of sets A such that for $A^m = A^m(A)$ and $\widetilde{X}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{im})' = (v'X_i)_{v \in \mathcal{V}(A^m)}$, $i = 1, \ldots, n$, the following conditions are satisfied:

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

and, in addition, one of the following conditions is satisfied:

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

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

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

PROPOSITION 3.1 (High-dimensional CLT for simple convex sets). *Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that conditions* (M.1′), (M.2′) *and* (E.1′) *are satisfied for every $A \in \mathcal{A}$. Then*

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

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

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

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 $\widetilde{X}_i = (v'X_i)_{v\in\mathcal{V}(A^m)}$ satisfy condition (E.1′) is the case where each X_i obeys a log-concave distribution. Recall that a Borel probability measure μ on $\mathbb{R}^p$ is *log-concave* if for any compact sets A_1, A_2 in $\mathbb{R}^p$ and $\lambda \in (0,1)$,

$$\mu(\lambda A_1 + (1-\lambda)A_2) \ge \mu(A_1)^{\lambda}\mu(A_2)^{1-\lambda},$$

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

COROLLARY 3.1 (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 $\mathrm{E}[X_i X_i']$ are bounded from below by a constant $k_1 > 0$ and from above by a constant $k_2 \ge k_1$ for every $i = 1, \dots, n$. Then*

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

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 $A = \bigcap_{q=1}^{Q} A_q$ 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 $A = \bigcap_{q=1}^{Q} A_q$ 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 $Q \le C_s^p \le p^s$, where C_s^p is the number of combinations of size s from p objects. Indeed, if we have a sparse representation $A = \bigcap_{q=1}^{Q} A_q$ 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 $Q \le C_s^p$.

EXAMPLE 3.1. The simplest example satisfying Definition 3.1 is a hyperrectangle as in (4), which is a 1-sparsely convex set. Another example is the set

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

for some unit vectors $v_k \in \mathbb{S}^{p-1}$ and coefficients a_k, $k = 1, \dots, m$. If the number of nonzero elements of each v_k does not exceed s, this A is an s-sparsely convex set. Yet another example is the set

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

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

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

for some coefficients $-\infty \le a_j \le b_j \le \infty$, $0 < c_{kl} \le \infty$, $j, k, l = 1, \dots, p$. This A is a 2-sparsely convex set. Finally, consider the set

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

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

As the proof of Proposition 3.2 below reveals, s-sparsely convex sets are closely related to simple convex sets. In particular, we can split any s-sparsely convex set $A \subset \mathbb{R}^p$ into $A \cap B$ and $A \cap B'$ for a cube $B = \{w \in \mathbb{R}^p : \max_{1 \le j \le p} |w_j| \le R\}$. Setting $R = pn^{5/2}$, it is easy to show that both $P(S_n^X \in A \cap B')$ and $P(S_n^Y \in A \cap B')$ are negligible. On the other hand, $A \cap B$ is a simple convex set with parameters $a = 1$ and d depending only s as long as $A \cap B$ contains a ball of radius $1/n$, and if $A \cap B$ does not contain such a ball, both $P(S_n^X \in A \cap B)$ and $P(S_n^Y \in A \cap B)$ are also negligible.

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

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

Then we have the following proposition.

PROPOSITION 3.2 (High-dimensional CLT for sparsely convex sets). *Suppose that conditions* (M.1″) *and* (M.2) *are satisfied. Then under* (E.1), *we have*

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

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

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

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 probabilities $\mathrm{P}(S_n^Y \in A)$ under weak conditions for hyperrectangles $A \in \mathcal{A}^{\mathrm{re}}$, simple convex sets $A \in \mathcal{A}^{\mathrm{si}}(a, d)$, or sparsely convex sets $A \in \mathcal{A}^{\mathrm{sp}}(s)$. 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. Hence, 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 data-dependent techniques. We consider here multiplier and empirical bootstrap methods (we refer to [32] for various versions of bootstraps).

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 $\bar{X} := (\bar{X}_1, \dots, \bar{X}_p)' := \mathbb{E}_n[X_i]$, and consider the normalized sum:

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

We are interested in bounding

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

for $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$, $\mathcal{A}^{\mathrm{sp}}(s)$, or $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$.

We begin with the case $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$. Let

$$\widehat{\Sigma} := n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})', \qquad \Sigma := n^{-1} \sum_{i=1}^n \mathrm{E}[X_i X_i'].$$

Observe that $\mathrm{E}[S_n^{eX}(S_n^{eX})' \mid X_1^n] = \widehat{\Sigma}$ and $\mathrm{E}[S_n^Y (S_n^Y)'] = \Sigma$. For $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$, define

$$\Delta_n(\mathcal{A}) := \sup_{A \in \mathcal{A}} \max_{v_1, v_2 \in \mathcal{V}(A^m(A))} \left|v_1'(\widehat{\Sigma} - \Sigma) v_2\right|.$$

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

THEOREM 4.1 (Abstract multiplier bootstrap theorem for simple convex sets). *Let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a, d)$ such that condition* (M.1′) *is satisfied for every $A \in \mathcal{A}$. Then for every constant $\overline{\Delta}_n > 0$, on the event $\Delta_n(\mathcal{A}) \le \overline{\Delta}_n$, we have*

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

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

REMARK 4.1 (Case of hyperrectangles). 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 $\overline{\Delta}_n > 0$, on the event $\Delta_{n,r} \le \overline{\Delta}_n$, we have

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

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

$$\Delta_{n,r} = \max_{1 \le j,k \le p} |\widehat{\Sigma}_{jk} - \Sigma_{jk}|,$$

where $\widehat{\Sigma}_{jk}$ and Σ_{jk} are the (j,k)th elements of $\widehat{\Sigma}$ and Σ, respectively.

Next, we derive more explicit bounds on $\rho_n^{\mathrm{MB}}(\mathcal{A})$ for $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$ under suitable moment conditions as in the previous section. We will consider sets $A \in \mathcal{A}^{\mathrm{si}}(a,d)$ that satisfy the following condition:

(S) The set $A^m = A^m(A)$ satisfies $\|v\|_0 \le s$ for all $v \in \mathcal{V}(A^m)$.

Condition (S) requires that the outward unit normal vectors to the hyperplanes forming the m-generated convex set $A^m = A^m(A)$ are sparse. Assuming that (S) is satisfied for all $A \in \mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$ helps to control $\Delta_n(\mathcal{A})$.

For $\alpha \in (0, e^{-1})$, define

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

Then we have the following proposition.

PROPOSITION 4.1 (Multiplier bootstrap for simple convex sets). *Let $\alpha \in (0, e^{-1})$ be a constant, and let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that conditions* (S) *and* (M.1$'$) *are satisfied for every $A \in \mathcal{A}$. In addition, suppose that condition* (M.2) *is satisfied. Then under* (E.1), *we have with probability at least $1-\alpha$,*

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

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

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

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

REMARK 4.2 (Bootstrap theorems in a.s. sense). Proposition 4.1 leads to the following multiplier bootstrap theorem in the a.s. sense. Suppose that $\mathcal{A}$ is a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ as in Proposition 4.1 and that (M.2) is 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 Proposition 4.1 with $\alpha = \alpha_n = n^{-1}(\log n)^{-2}$, together with the

Borel–Cantelli lemma (note that $\sum_{n=4}^{\infty} n^{-1}(\log n)^{-2} < \infty$), we have with probability one

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

and it is routine to verify that $D_n^{(1)}(\alpha_n) = o(1)$ if $B_n^2 \log^7(pn) = o(n)$, and $D_{n,q}^{(2)}(\alpha_n) = o(1)$ if $B_n^2(\log^3(pn))\log^{4/q} n = o(n^{1-4/q})$. Similar conclusions also follow from other propositions and corollaries below dealing with different classes of sets and approximations based on multiplier and empirical bootstraps.

When each X_i obeys a log-concave distribution, we have the following corollary analogous to Corollary 3.1. In this case, instead of condition (S), we will assume that $\mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a,d)$ is such that the cardinality of the set $\bigcup_{A \in \mathcal{A}} \mathcal{V}(A^m(A))$ is at most $(pn)^d$.

COROLLARY 4.1 (Multiplier bootstrap for simple convex sets with log-concave distributions). *Let $\alpha \in (0, e^{-1})$ be a constant, and let $\mathcal{A}$ be a subclass of $\mathcal{A}^{\mathrm{si}}(a,d)$ such that the cardinality of the set $\bigcup_{A \in \mathcal{A}} \mathcal{V}(A^m(A))$ is at most $(pn)^d$. 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 every $i = 1, \dots, n$. Then with probability at least $1-\alpha$,*

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

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

When $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$, we have the following corollary.

COROLLARY 4.2 (Multiplier bootstrap for hyperrectangles). *Let $\alpha \in (0, e^{-1})$ be a constant, and suppose that conditions* (M.1) *and* (M.2) *are satisfied. Then under* (E.1), *we have with probability at least $1-\alpha$,*

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

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

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

where the constant C depends only on b and q.

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

PROPOSITION 4.2 (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* $1-\alpha$,

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

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

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

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

4.2. *Empirical bootstrap*. Here, we consider the empirical bootstrap. For brevity, we only consider the case $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$. 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 $\bar{X} = \mathbb{E}_n[X_i]$. Consider the normalized sum:

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

We are interested in bounding

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

for $\mathcal{A} = \mathcal{A}^{\mathrm{re}}$. To state the bound, define

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

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

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

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

which are empirical analogs of $M_{n,X}(\phi)$ and $M_{n,Y}(\phi)$, respectively. Let

$$\widehat{M}_n(\phi) := \widehat{M}_{n,X}(\phi) + \widehat{M}_{n,Y}(\phi).$$

We have the following theorem.

THEOREM 4.2 (Abstract empirical bootstrap theorem). *For arbitrary positive constants* b, $\overline{L}_n$ *and* $\overline{M}_n$, *the inequality*

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

holds on the event

$$\{\mathbb{E}_n[(X_{ij}-\bar{X}_j)^2] \geq b \text{ for all } j=1,\dots,p\} \cap \{\widehat{L}_n \leq \overline{L}_n\} \cap \{\widehat{M}_n(\phi_n) \leq \overline{M}_n\},$$

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 for the multiplier bootstrap case, we next derive explicit bounds on $\rho_n^{\mathrm{EB}}(\mathcal{A}^{\mathrm{re}})$ under suitable moment conditions.

PROPOSITION 4.3 (Empirical bootstrap for hyperrectangles). *Let* $\alpha \in (0, e^{-1})$ *be a constant*, *and suppose that conditions* (M.1) *and* (M.2) *are satisfied*. *In addition*, *suppose that* $\log(1/\alpha) \leq K \log(pn)$ *for some constant* K. *Then under* (E.1), *we have with probability at least* $1-\alpha$,

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

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

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

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

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

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

where the random vectors $Y_1, \dots, Y_n$ are assumed to be independent of the random vectors $X_1, \dots, X_n$, and recall that $M_n(\phi) := M_{n,X}(\phi) + M_{n,Y}(\phi)$ for $\phi \geq 1$. The lemma below provides a bound on ϱ_n.

LEMMA 5.1 (Key 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* $\phi \geq 1$:

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

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

Lemma 5.1 has an immediate corollary. Indeed, define

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

where $\mathcal{A}^{\mathrm{re}}$ is the class of all hyperrectangles in $\mathbb{R}^p$. Then we have the following.

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 $\phi \geq 1$:*

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

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 inequality due to Nazarov [30].

LEMMA A.1 (Nazarov's inequality, [30]). *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 $a > 0$,*

$$\mathrm{P}(Y \leq y + a) - \mathrm{P}(Y \leq y) \leq Ca\sqrt{\log p},$$

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 [20] for the case of max hyperrectangles. However, the former inequality allows for more general hyperrectangles than the latter. The difference in sharpness for the case of max-hyperrectangles arises due to dimension-dependence $\sqrt{\log p}$, in particular the term $\sqrt{\log p}$ can be much larger than $\mathrm{E}[\max_{1 \leq j \leq p} Y_j]$. This also makes the anti-concentration bound in [20] 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 [20] for classes of hyperrectangles other than max hyperrectangles.

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

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

and similarly

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

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 [25], whose proof the authors credit to Nazarov [30]. □

We will use another anti-concentration inequality by [30] in the proofs for Sections 3 and 4, 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 $Q \subset \mathbb{R}^p$, and every $h_1, h_2 > 0$,*

$$\frac{\gamma_A(Q^{h_1} \setminus Q^{-h_2})}{h_1 + h_2} \leq C\sqrt{\|A\|_{\mathrm{HS}}},$$

where $\|A\|_{\mathrm{HS}}$ is the Hilbert–Schmidt norm of A, $Q^h = \{x \in \mathbb{R}^p : \rho(x, Q) \leq h\}$, $Q^{-h} = \{x \in \mathbb{R}^p : B(x, h) \subset Q\}$, $B(x, h) = \{y \in \mathbb{R}^p : \|y - x\| \leq h\}$, and $\rho(x, Q) = \inf_{y \in Q} \|y - x\|$.

PROOF. It is proven in [30] that for every convex set $Q \subset \mathbb{R}^p$ and every $h > 0$,

$$\frac{\gamma_A(Q^h \setminus Q)}{h} \leq C\sqrt{\|A\|_{\mathrm{HS}}}.$$

Therefore, the asserted claim follows from the arguments in Proposition 2.5 of [16] or in Section 1.3 of [8]. □

APPENDIX B: PROOF FOR SECTION 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 nondecreasing functions, and let ξ_i, $i = 1, 2$ be independent real-valued random variables. Then*

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

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

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

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

PROOF. Inequality (18) is Chebyshev's association inequality; see Theorem 2.14 in [12]. Moreover, since ξ_1 and ξ_2 are independent, (19) follows from (20). In turn, (20) follows from

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

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

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

which follows from monotonicity of φ_1 and φ_2. $\square$

Proof of Lemma 5.1. The proof relies on a Slepian–Stein method developed in [17]. 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 $S_n^W := n^{-1/2}\sum_{i=1}^n W_i$. Then $\mathrm{P}(S_n^Y \le y) = \mathrm{P}(S_n^W \le y)$, so that

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

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

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

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

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

Pick 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 $t \le 0$ and $g_0(t) = 0$ for $t \ge 1$. Define $g(t) := g_0(\phi t)$, $t \in \mathbb{R}$, and

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

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 properties established in Lemmas A.5 and A.6 of [17]: for every $j, k, l = 1, \dots, p$, there exists a function $U_{jkl}(w)$ such that

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

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

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

where the inequalities (22) and (23) hold for all $w \in \mathbb{R}^p$, and inequality (24) holds for all $w, \widetilde{w} \in \mathbb{R}^p$ with $\max_{1\le j\le p} |\widetilde{w}_j|\beta \le 1$ (formally, [17] only considered the case where $y = (0, \dots, 0)'$ but the extension to $y \in \mathbb{R}^p$ is trivial). Moreover, define the functions

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

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

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

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

where

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

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

Step 1. Define the Slepian interpolant

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

where

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

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

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

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

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

where

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

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

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

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

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

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

To this end, let

$$\chi_i := 1\Bigl\{\max_{1\le j\le p}|X_{ij}| \vee |Y_{ij}| \vee |W_{ij}| \le \sqrt{n}/(4\beta)\Bigr\}, \qquad i = 1,\dots,n$$

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

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

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

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

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

$$\lesssim \phi\beta^2 \sum_{i=1}^{n} \int_0^1 \mathrm{E}\Big[(1-\chi_i) \max_{1\le j,k,l\le p} |\dot{Z}_{ij} Z_{ik} Z_{il}|\Big]\, dt \tag{28}$$

$$\lesssim \frac{\phi\beta^2}{n^{3/2}} \sum_{i=1}^{n} \int_0^1 \omega(t) \mathrm{E}\Big[(1-\chi_i) \max_{1\le j\le p} |X_{ij}|^3 \vee |Y_{ij}|^3 \vee |W_{ij}|^3\Big]\, dt,$$

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

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

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

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

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

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

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 $h(Z^{(i)}, 1) = 0$ and $\chi_i = 1$, so that

$$\chi_i |m_{jkl}\big(Z^{(i)} + \tau Z_i\big)| = h\big(Z^{(i)}, 1\big) \chi_i |m_{jkl}\big(Z^{(i)} + \tau Z_i\big)|. \tag{29}$$

Indeed if $\chi_i = 1$, then $\max_{1\le j\le p} |Z_{ij}| \le 3/(4\beta) < 1/\beta$, and so when $h(Z^{(i)}, 1) = 0$ 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,

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

$$\lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \int_0^1 \mathrm{E}\big[\chi_i h\big(Z^{(i)}, 1\big) U_{jkl}\big(Z^{(i)} + \tau Z_i\big) |\dot{Z}_{ij} Z_{ik} Z_{il}|\big]\, d\tau\, dt \tag{30}$$

$$\lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \int_0^1 \mathrm{E}\big[\chi_i h\big(Z^{(i)}, 1\big) U_{jkl}\big(Z^{(i)}\big) |\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, d\tau \, dt$$

$$\lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}\big[h\big(Z^{(i)}, 1\big) U_{jkl}\big(Z^{(i)}\big)\big] \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt,$$

where the second inequality follows from (22) and (29), the third inequality from (24), and the fourth inequality from the independence of $Z^{(i)}$ from $\dot{Z}_{ij} Z_{ik} Z_{il}$. Then we split the integral in (30) by inserting $\chi_i + (1 - \chi_i)$ under the first expectation sign. We have

$$\sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}\big[(1 - \chi_i) h\big(Z^{(i)}, 1\big) U_{jkl}\big(Z^{(i)}\big)\big] \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt$$

$$\lesssim \phi\beta^2 \sum_{i=1}^{n} \int_0^1 \mathrm{E}[1 - \chi_i] \mathrm{E}\Big[\max_{1 \le j,k,l \le p} |\dot{Z}_{ij} Z_{ik} Z_{il}|\Big] \, dt$$

$$\lesssim M_n(\phi)\phi\beta^2 / n^{1/2},$$

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 $h(Z^{(i)}, 1) = 0$ whenever $h(Z, 2) = 0$ and $\chi_i = 1$ (which follows from the same argument as before), so that

$$\chi_i h\big(Z^{(i)}, 1\big) = \chi_i h\big(Z^{(i)}, 1\big) h(Z, 2),$$

we have

$$\sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}\big[\chi_i h\big(Z^{(i)}, 1\big) U_{jkl}\big(Z^{(i)}\big)\big] \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt$$

$$\lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}\big[\chi_i h\big(Z^{(i)}, 1\big) U_{jkl}(Z)\big] \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt$$

$$\lesssim \sum_{j,k,l=1}^{p} \sum_{i=1}^{n} \int_0^1 \mathrm{E}\big[h(Z, 2) U_{jkl}(Z)\big] \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt \tag{31}$$

$$= \sum_{j,k,l=1}^{p} \int_0^1 \mathrm{E}\big[h(Z, 2) U_{jkl}(Z)\big] \sum_{i=1}^{n} \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt$$

$$\lesssim \phi\beta^2 \int_0^1 \mathrm{E}\big[h(Z, 2)\big] \max_{1 \le j,k,l \le p} \sum_{i=1}^{n} \mathrm{E}\big[|\dot{Z}_{ij} Z_{ik} Z_{il}|\big] \, dt.$$

To bound (31), observe that

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

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

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

Meanwhile, observe that

$$\mathrm{E}\big[h(Z,2)\big] = \mathrm{P}(Z \le \overline{I}) - \mathrm{P}(Z \le \underline{I}),$$

where

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\sqrt{tv} X_i + \sqrt{t(1-v)} Y_i + \sqrt{1-t} W_i\big) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\sqrt{tv} X_i + \sqrt{1-tv} Y_i\big),$$

and $\underline{I} = y - \phi^{-1} - 2\beta^{-1}$, $\overline{I} = y + \phi^{-1} + 2\beta^{-1}$; 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,

$$\mathrm{P}(Z \le \overline{I}) \le \mathrm{P}\big(S_n^Y \le \overline{I}\big) + \varrho_n, \qquad \mathrm{P}(Z \le \underline{I}) \ge \mathrm{P}\big(S_n^Y \le \underline{I}\big) - \varrho_n,$$

and by Lemma A.1,

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

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

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

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

$$(31) \lesssim \frac{\phi\beta^2 L_n}{n^{1/2}} \big(\varrho_n + \phi^{-1} \log^{1/2} p\big) \lesssim \frac{\phi^2 \log^2 p}{n^{1/2}} \big(\phi L_n \varrho_n + L_n \log^{1/2} p\big),$$

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

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

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

Then we have

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

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

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

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 hyperrectangle

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

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

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

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 $\phi_n \ge 2$; 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

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

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 Proposition 2.1, we shall verify the following elementary inequality.

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

PROOF. Observe that

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

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

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

which leads to

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

completing the proof. $\square$

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

$$\frac{B_n^2 \log^7(pn)}{n} \leq c := \min\big\{(c_1/2)^3, (K_2/2)^6\big\}, \tag{32}$$

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

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

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

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

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

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). Set $\overline{L}_n := B_n$. 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. In addition, since each Y_{ij} is Gaussian and $\mathrm{E}[Y_{ij}^2] = \mathrm{E}[X_{ij}^2]$, $\|Y_{ij}\|_{\psi_1} \leq C_1 B_n$ for all i and j and some universal constant $C_1 > 0$. Hence, by Lemma 2.2.2 in [42], we have for some universal constant $C_2 > 0$, $\|\max_{1\leq j\leq p} X_{ij}\|_{\psi_1} \leq C_2 B_n \log p$ and $\|\max_{1\leq j\leq p} Y_{ij}\|_{\psi_1} \leq C_2 B_n \log p$. Together with Markov's inequality, this implies that for every $t > 0$,

$$\mathrm{P}\Big(\max_{1\leq j\leq p} |X_{ij}| > t\Big) \leq 2\exp\Big(-\frac{t}{C_2 B_n \log p}\Big).$$

Applying Lemma C.1, we have

$$M_{n,X}(\phi_n) \lesssim \big(\sqrt{n}/(\phi_n \log p) + B_n \log p\big)^3 \exp\Big(-\frac{\sqrt{n}}{4C_2\phi_n B_n \log^2 p}\Big).$$

Here,

$$\begin{aligned}\frac{\sqrt{n}}{4C_2\phi_n B_n \log^2 p} &= \frac{c_1 n^{1/3}}{B_n^{2/3}\log^{4/3} p} \qquad \Big(c_1 := \frac{1}{4K_2C_2}\Big)\\ &\geq c_1 c^{-1/3}\log(pn) \geq 2\log(pn) \qquad \big(\text{by (32)}\big).\end{aligned}$$

Moreover, by (32) 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 $(\sqrt{n}/(\phi_n \log p) + B_n \log p)^3 \lesssim n^{3/2}$, which implies that

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

The same reasoning also gives $M_{n,Y}(\phi_n) \lesssim n^{-1/2}$. The conclusion of the proposition in this case now 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 (32), we may assume that

$$\frac{B_n \log^{3/2} p}{n^{1/2-1/q}} \leq (K_2/2)^{3/2}. \tag{33}$$

Set

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

Then $L_n \leq B_n \leq \overline{L}_n$. As the map $x \mapsto x^{1/3}$ is sub-linear, $\{n^{-1}\overline{L}_n^2\log^7 p\}^{1/6} \leq D_n^{(1)} + D_{n,q}^{(2)} \leq K_2$, so that $\phi_n^{-1} = K_2^{-1}\{n^{-1}\overline{L}_n^2\log^4 p\}^{1/6} \leq 1$.

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

$$\begin{aligned} M_{n,X}(\phi_n)/\overline{L}_n &\lesssim \frac{B_n^{q/3+2} (\log p)^{q/2-3/2}}{n^{q/6+1/6-2/q}} \cdot \frac{n^{1/2-2/q} \log^{1/2} p}{B_n^2} \\ &\lesssim \frac{1}{\log p} \left(\frac{B_n^2 \log^3 p}{n^{1-2/q}} \right)^{q/6} \lesssim D_{n,q}^{(2)}. \end{aligned}$$

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 Proposition 3.1. Here, C denotes a generic 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 $A \in \mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$. Let $A^m = A^m(A)$ be an approximating m-generated convex set as in condition (C). By assumption, $A^m \subset A \subset A^{m,\epsilon}$, so that by letting

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

we have $\mathrm{P}(S_n^X \in A) \le \mathrm{P}(S_n^X \in A^{m,\epsilon}) \le \mathrm{P}(S_n^Y \in A^{m,\epsilon}) + \overline{\rho}$. Here, observe that $(v' S_n^Y)_{v \in \mathcal{V}(A^m)}$ is a Gaussian random vector with dimension $\mathrm{Card}(\mathcal{V}(A^m)) = m \le (pn)^d$ such that, by condition (M.1′), the variance of each coordinate is bounded from below by b. Hence, by Lemma A.1, we have

$$\begin{aligned} \mathrm{P}(S_n^Y \in A^{m,\epsilon}) &= \mathrm{P}\{v' S_n^Y \le \mathcal{S}_{A^m}(v) + \epsilon \text{ for all } v \in \mathcal{V}(A^m)\} \\ &\le \mathrm{P}\{v' S_n^Y \le \mathcal{S}_{A^m}(v) \text{ for all } v \in \mathcal{V}(A^m)\} + C\epsilon \log^{1/2} m \\ &= \mathrm{P}(S_n^Y \in A^m) + C\epsilon \log^{1/2}(pn), \end{aligned}$$

so that

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

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

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

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

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

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

Proof of Corollary 3.1. Since X_i is a centered random vector with a log-concave distribution in $\mathbb{R}^p$, Borell's inequality (see [11], 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 [28], 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 $A \in \mathcal{A}^{\mathrm{si}}(a,d)$ 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 $A \in \mathcal{A}^{\mathrm{si}}(a,d)$ 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 Proposition 3.1.

Proof of Proposition 3.2. 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.

Let $R := pn^{5/2}$ and $V^R := \{w \in \mathbb{R}^p : \max_{1 \leq j \leq p} |w_j| > R\}$. Fix any $A \in \mathcal{A}^{\mathrm{sp}}(s)$. Then $A = \check{A} \cup (A \cap V^R)$ for some s-sparsely convex set $\check{A} \subset \mathbb{R}^p$ such that $\sup_{w \in \check{A}} \max_{1 \leq j \leq p} |w_j| \leq R$. Now observe that by Markov's inequality,

$$\begin{aligned} \mathrm{P}\Big(\max_{i,j} |X_{ij}| > pn^2\Big) &\leq \frac{\mathrm{E}[\max_{i,j} |X_{ij}|]}{pn^2} \leq \frac{\mathrm{E}[\sum_{i,j} |X_{ij}|]}{pn^2} \\ &\leq \max_{i,j} \mathrm{E}\big[|X_{ij}|\big]/n \leq C B_n/n \leq C/n^{1/2}, \end{aligned}$$

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

$$\mathrm{P}(S_n^X \in V^R) \leq C/n^{1/2},$$

and similarly,

$$\mathrm{P}(S_n^Y \in V^R) \leq C/n^{1/2}.$$

So,

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

Therefore, it suffices to consider the case where the sets $A \in \mathcal{A}^{\mathrm{sp}}(s)$ are such that

$$\sup_{w \in A} \max_{1 \le j \le p} |w_j| \le R. \tag{34}$$

Further, let $\varepsilon = n^{-1}$, and define $\mathcal{A}_1^{\mathrm{sp}}(s)$ as the class of all sets $A \in \mathcal{A}^{\mathrm{sp}}(s)$ satisfying (34) and containing a ball with radius ε and center at, say, w_A. Also define $\mathcal{A}_2^{\mathrm{sp}}(s)$ as the class of all sets $A \subset \mathcal{A}^{\mathrm{sp}}(s)$ satisfying (34) and containing no ball of radius ε. We bound $\rho_n(\mathcal{A}_1^{\mathrm{sp}}(s))$ and $\rho_n(\mathcal{A}_2^{\mathrm{sp}}(s))$ separately in two steps. In both cases, we rely on the following lemma, whose proof is given after the proof of this proposition.

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

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

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

$$\|v\|_0 \le s \qquad \text{for all } v \in \mathcal{V}(A^m). \tag{35}$$

Therefore, since $Q \le p^s$, if $R \le (pn)^{d_0}$ and $\mu \le (pn)^{d_0}$ for some constant $d_0 \ge 1$, then the set A satisfies condition (C) *with $a = 1$ and d depending only on s and d_0, and the approximating m-generated convex set A^m satisfying* (35).

Step 1. Here, we bound $\rho_n(\mathcal{A}_1^{\mathrm{sp}}(s))$. Pick any s-sparsely convex set $A \in \mathcal{A}_1^{\mathrm{sp}}(s)$ with a sparse representation $A = \bigcap_{q=1}^{Q} A_q$ for some $Q \le p^s$. Below we verify conditions (C), (M.1′), (M.2′) and (E.1′) [or (E.2′)] 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 $\|w\| \le 2p^{1/2}R$ for all $w \in B$, and has a sparse representation $B = \bigcap_{q=1}^{Q} B_q$ 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 $Q \le p^s$, we see that the set A satisfies condition (C) with $a = 1$ and d depending only on s, and an approximating m-generated convex set A^m such that $\|v\|_0 \le s$ for all $v \in \mathcal{V}(A^m)$.

Further, since we have $\|v\|_0 \le s$ for all $v \in \mathcal{V}(A^m)$, the fact that the set A satisfies condition (M.1′) follows immediately from (M.1″).

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

$$n^{-1}\sum_{i=1}^n \mathrm{E}\big[|v'X_i|^{2+k}\big] \le n^{-1}\sum_{i=1}^n \mathrm{E}\bigg[\bigg(\sum_{j\in J(v)} |X_{ij}|\bigg)^{2+k}\bigg]$$
$$\le s^{1+k} n^{-1}\sum_{i=1}^n \mathrm{E}\bigg[\sum_{j\in J(v)} |X_{ij}|^{2+k}\bigg] \le s^{2+k} B_n^k \le \big(B_n'\big)^k$$

for $k = 1$ or 2, where $B_n' = s^3 B_n$, so that the set A satisfies condition (M.2$'$) with B_n replaced by $s^3 B_n$.

Finally, we verify that the set A satisfies condition (E.1$'$) when (E.1) is satisfied, or (E.2$'$) when (E.2) is satisfied. When (E.1) is satisfied, we have $\|X_{ij}\|_{\psi_1} \le B_n$, so that $\|v'X_i\|_{\psi_1} \le \sum_{j\in J(v)} \|X_{ij}\|_{\psi_1} \le sB_n$ showing that the set A satisfies (E.1$'$) with B_n replaced by sB_n.

When (E.2) is satisfied, as $\mathrm{E}[\max_{v\in\mathcal{V}(A^m)} |v'X_i|^q] \le s^q \mathrm{E}[\max_{1\le j\le p} |X_{ij}|^q]$, the set A satisfies (E.2$'$) with B_n replaced by sB_n.

Thus, all sets $A \in \mathcal{A}_1^{\mathrm{sp}}(s)$ satisfy conditions (C), (M.1$'$), (M.2$'$) and (E.1$'$) [or (E.2$'$)], and so applying Proposition 3.1 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 2. Here, we bound $\rho_n(\mathcal{A}_2^{\mathrm{sp}}(s))$. Fix any s-sparsely convex set $A \in \mathcal{A}_2^{\mathrm{sp}}(s)$ with a sparse representation $A = \bigcap_{q=1}^Q A_q$ for some $Q \le p^s$. We consider two cases separately. First, suppose that at least one A_q does not contain a ball with radius ε. Then under condition (M.1$''$), 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 [24], Theorem 1.3) implies that

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

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 $CB_n/n^{1/2}$, and so is absolute value of 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 $m \le (pn)^d$ 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 $A^0 = \bigcap_{q=1}^Q A_q^{m,1/n}$ 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 Step 1 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 Proposition 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 $\bigcap_{q=1}^Q A_q^{m,-\varepsilon}$ is empty because $\bigcap_{q=1}^Q A_q^m \subset A$ and A contains no ball with radius ε. This implies that $\mathrm{P}(S_n^Y \in A^0) \le C(\log^{1/2}(pn))/n$ 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 Proposition 3.2.

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

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

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 $P_3 \subset P_4$,

$$d_{\mathrm{BM}}(P_1 \cap P_3, P_2 \cap P_4) \le d_{\mathrm{BM}}(P_1, P_2) \vee d_{\mathrm{BM}}(P_3, P_4). \tag{36}$$

Let $A = \bigcap_{q=1}^Q A_q$ 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 $s_q \le s$ 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 were 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 $v \in \mathcal{V}(P_q)$, nonzero elements of v correspond to some of the main components of A_q. Since we need at most s_q vertices to form a facet of the polytope P_q, the polytope P_q has

$$m_q \le \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s_q^2} \le \left(\gamma\sqrt{\frac{\mu+1}{\epsilon}}\log\frac{1}{\epsilon}\right)^{s^2} \tag{37}$$

facets. 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 nonzero elements. Hence, $d_{\mathrm{BM}}(P_q, A_q) \le \epsilon$, which, together with (36), implies that

$$d_{\mathrm{BM}}\left(\bigcap_{q=1}^Q P_q, \bigcap_{q=1}^Q A_q\right) \le \epsilon.$$

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

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

where the last assertion follows from the assumption that $\sup_{w\in A}\|w\| \le R$. Since A^m is an m-generated convex set with $m \le \sum_{q=1}^{Q} m_q$, the first claim of the lemma now follows from (37). The second claim (35) holds by construction of A^m, and the final claim is trivial. $\square$

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 $p \ge 2$. Define $Z := \max_{1\le j\le p} |\sum_{i=1}^n X_{ij}|$, $M := \max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|$ and $\sigma^2 := \max_{1\le j\le p}\sum_{i=1}^n \mathrm{E}[X_{ij}^2]$. Then*

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

where K is a universal constant.

PROOF. See Lemma 8 in [20]. $\square$

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

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

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

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

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

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 [22]. $\square$

LEMMA E.3. *Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^p$ with $p \ge 2$ such that $X_{ij} \ge 0$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$. Define $Z := \max_{1\le j\le p}\sum_{i=1}^n X_{ij}$ and $M := \max_{1\le i\le n}\max_{1\le j\le p} X_{ij}$. Then*

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

where K is a universal constant.

PROOF. See Lemma 9 in [20]. □

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 [27].

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. By homogeneity, we may assume that $B = 1$. Then by Theorem 10 in [27], 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 +

$\eta)8\mathrm{E}[M] \le t/4$, since otherwise we can make the lemma trivial by setting K or K' large enough. Take

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

Define

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

Then

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

Observe that

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

so that by the Hoffmann–Jørgensen inequality (see [26], 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 $C_1, C_2, \ldots$ are constants that depend only on η, β, s.

Case (i). By Theorem 6.21 in [26] (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 [26] (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 $t \to \infty$. $\square$

E.2. Proofs for Section 4.

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 $A \in \mathcal{A} \subset \mathcal{A}^{\mathrm{si}}(a, d)$. Let $A^m = A^m(A)$ be an approximating m-generated convex set as in (C). By assumption, $A^m \subset A \subset A^{m,\epsilon}$. Let

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

As in the proof of Proposition 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

$$\rho_n^{\mathrm{MB}}(\mathcal{A}^{\mathrm{re}}) \le C\overline{\Delta}_n^{1/3}\log^{2/3} p \tag{38}$$

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

To this end, we first show that

$$\varrho_n^{\mathrm{MB}} := \sup_{y\in\mathbb{R}^p}|\mathrm{P}(S_n^{eX} \le y \mid X_1^n) - \mathrm{P}(S_n^Y \le y)| \le C\Delta_{n,r}^{1/3}\log^{2/3} p. \tag{39}$$

To show (39), fix any $y = (y_1, \dots, y_p)' \in \mathbb{R}^p$. As in the proof of Lemma 5.1, for $\beta > 0$, define

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

Note that conditional on X_1^n, S_n^{eX} is a centered Gaussian random vector with covariance matrix $\widehat{\Sigma}$. Then a small modification of the proof of Theorem 1 in [20] implies that for every $g \in C^2(\mathbb{R})$ with $\|g'\|_\infty \vee \|g''\|_\infty < \infty$, we have

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

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

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

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 (39). Finally, (38) follows from the fact that the inequality $\varrho_n^{\mathrm{MB}} \le C\overline{\Delta}_n^{1/3}\log^{2/3} p$ holds on the event $\Delta_{n,r} \le \overline{\Delta}_n$, and applying the same argument as that used in the proof of Corollary 5.1.

Proof of Proposition 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

$$B_n^2 \log^5(pn) \log^2(1/\alpha) \leq n \tag{40}$$

since otherwise the assertions are trivial.

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

We begin with noting that since (S) holds for all $A \in \mathcal{A}$, $\Delta_n(\mathcal{A}) \leq C \Delta_{n,r}$ where $\Delta_{n,r} = \max_{1 \leq j,k \leq p} |\widehat{\Sigma}_{jk} - \Sigma_{jk}|$. As $\widehat{\Sigma} - \Sigma = n^{-1} \sum_{i=1}^n (X_i X_i' - \mathrm{E}[X_i X_i']) - \bar{X}\bar{X}'$, we have $\Delta_{n,r} \leq \Delta_{n,r}^{(1)} + \{\Delta_{n,r}^{(2)}\}^2$, where

$$\Delta_{n,r}^{(1)} := \max_{1 \leq j,k \leq p} \left| n^{-1} \sum_{i=1}^n \bigl(X_{ij} X_{ik} - \mathrm{E}[X_{ij} X_{ik}]\bigr) \right|, \qquad \Delta_{n,r}^{(2)} := \max_{1 \leq j \leq p} |\bar{X}_j|.$$

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

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

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

In addition, by (E.1),

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

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

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

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

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

where the last inequality follows from (40). Applying Lemma E.2(i) with $\beta = 1/2$ and $\eta = 1$, we conclude that for every $t > 0$,

$$\mathrm{P}\big(\Delta_{n,r}^{(1)} > C\{n^{-1}B_n^2 \log(pn)\}^{1/2} + t\big)$$
$$\le \exp\{-nt^2/(3B_n^2)\} + 3\exp\{-c\sqrt{nt}/\big(B_n \log(pn)\big)\}.$$

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 of this inequality is bounded by

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

where the last inequality follows from (40). Therefore,

$$\mathrm{P}\big(\{\Delta_{n,r}^{(1)} \log^2(pn)\}^{1/3} > CD_n^{(1)}(\alpha)\big) \le \alpha/2.$$

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 proposition under (E.1).

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

$$\mathrm{E}\big[M_n^{q/2}\big] \le C\Big\{\mathrm{E}\Big[\max_{i,j,k}|X_{ij}X_{ik}|^{q/2}\Big] + \max_{i,j,k}\big(\mathrm{E}\big[|X_{ij}X_{ik}|\big]\big)^{q/2}\Big\}$$
$$\le C\Big\{\mathrm{E}\Big[\max_{i,j,k}|X_{ij}X_{ik}|^{q/2}\Big]\Big\} = C\mathrm{E}\Big[\max_{i,j}|X_{ij}|^q\Big] \le CnB_n^q,$$

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

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

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

$$\mathrm{P}\big\{\Delta_{n,r}^{(1)} > C\big\{\big(n^{-1}B_n^2\log p\big)^{1/2} + n^{-1+2/q}B_n^2\log p\big\} + t\big\}$$
$$\le \exp\{-nt^2/(3B_n^2)\} + ct^{-q/2}n^{1-q/2}B_n^q.$$

Choosing

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

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

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

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 proposition under (E.2).

Proof of Corollary 4.1. 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$'$) for all $A \in \mathcal{A}$ and derive a suitable bound on $\Delta_n(\mathcal{A})$. Condition (M.1$'$) for all $A \in \mathcal{A}$ 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 $\|v'X_i\|_{\psi_1} \leq C(\mathrm{E}[(v'X_i)^2])^{1/2} \leq C$ for all $v \in \mathbb{R}^p$ with $\|v\| = 1$ (see the proof of Corollary 3.1). For all $i = 1, \dots, n$, let $\check{X}_i$ be a random vector whose elements are given by $v'X_i$, $v \in \bigcup_{A \in \mathcal{A}} \mathcal{V}(A^m(A))$; the dimension of $\check{X}_i$, denoted by $\check{p}$, is at most $(pn)^d$, and $\|\check{X}_{ij}\|_{\psi_1} \leq C$ for all $j = 1, \dots, \check{p}$. Then $\Delta_n(\mathcal{A})$ coincides with $\Delta_{n,r}$ with X_i replaced by $\check{X}_i$, that is,

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

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

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

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

Proof of Corollary 4.2. Any hyperrectangle $A \in \mathcal{A}^{\mathrm{re}}$ satisfies conditions (C) and (S) with $a = 0$, $d = 1$, and $s = 1$. In addition, it follows from (M.1) that any hyperrectangle $A \in \mathcal{A}^{\mathrm{re}}$ satisfies (M.1$'$). Therefore, the asserted claims follow from Proposition 4.1.

Proof of Proposition 4.2. In this proof, let C be a positive constant depending only on b and s under (E.1), and on b, q and s under (E.2); the value of C may change from place to place. Moreover, without loss of generality, we will assume that

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

since otherwise the assertions are trivial.

Let $\Delta_{n,r} := \max_{1 \leq j,k \leq p} |\widehat{\Sigma}_{jk} - \Sigma_{jk}|$, and

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

Then by the proof of Proposition 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

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

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

As in the proof of Proposition 3.2, let $R = pn^{5/2}$ and $V^R = \{w \in \mathbb{R}^p : \max_{1\leq j\leq p}|w_j| > R\}$. Fix any $A \in \mathcal{A}^{\mathrm{sp}}(s)$. Then $A = \check{A} \cup (A \cap V^R)$ for some s-sparsely convex set $\check{A}$ with $\sup_{w\in\check{A}}\max_{1\leq j\leq p}|w_j| \leq R$. As in Proposition 3.2, $\mathrm{P}(S_n^Y \in V^R) \leq C/n^{1/2}$. Moreover, conditional on X_1^n, S_{nj}^{eX} is Gaussian with mean zero and variance $\mathbb{E}_n[(X_{ij} - \bar{X}_j)^2] = \widehat{\Sigma}_{jj}$, so that

$$\begin{aligned}
\mathrm{P}\big(S_n^{eX} \in V^R \mid X_1^n\big) &= \mathrm{P}\Big(\max_{1\leq j\leq p}\big|S_{nj}^{eX}\big| > R \mid X_1^n\Big) \\
&\leq \frac{\mathrm{E}[\max_{1\leq j\leq p}|S_{nj}^{eX}| \mid X_1^n]}{R} \\
&\leq \frac{C(\log p)^{1/2}\max_{1\leq j\leq p}\widehat{\Sigma}_{jj}^{1/2}}{R},
\end{aligned}$$

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

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

so that it suffices to consider the case where the sets $A \in \mathcal{A}^{\mathrm{sp}}(s)$ are such that $\sup_{w\in A}\max_{1\leq j\leq p}|w_j| \leq R$.

Further, let $\varepsilon = n^{-1}$, and define the subclasses $\mathcal{A}_1^{\mathrm{sp}}(s)$ and $\mathcal{A}_2^{\mathrm{sp}}(s)$ of $\mathcal{A}^{\mathrm{sp}}(s)$ as in the proof of Proposition 3.2. For all $A \in \mathcal{A}_1^{\mathrm{sp}}(s)$, we can verify conditions (C), (S), and (M.1$'$) as in the proof of Proposition 3.2 [where (S) is verified implicitly]. Therefore, by Proposition 4.1 applied with $\alpha/2$ instead of α, the bounds (14) and (15) with $\rho_n^{\mathrm{MB}}(\mathcal{A}^{\mathrm{sp}}(s))$ replaced by $\rho_n^{\mathrm{MB}}(\mathcal{A}_1^{\mathrm{sp}}(s))$ hold with probability at least $1 - \alpha/2$. Hence, it remains to bound $\rho_n^{\mathrm{MB}}(\mathcal{A}_2^{\mathrm{sp}}(s))$.

Fix any $A \in \mathcal{A}_2^{\mathrm{sp}}(s)$ with a sparse representation $A = \bigcap_{q=1}^{Q} A_q$ for some $Q \leq p^s$. As in the proof of Proposition 3.2, we separately consider two cases. First, suppose that at least one of A_q does not contain a ball of radius ε; then by condition (M.1$''$) and Lemma A.2, $\mathrm{P}(S_n^Y \in A_q) \leq C\varepsilon$. 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 $\Delta_{n,r} \leq C_1\overline{\Delta}_n$, $\mathrm{P}(S_n^{eX} \in A_q \mid X_1^n) \leq C\varepsilon$ since $C_1\overline{\Delta}_n \leq b/2$. Since $A \subset A_q$, we conclude that on the event $\Delta_{n,r} \leq C_1\overline{\Delta}_n$, $|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| \leq C\varepsilon = C/n$.

Second, suppose that each A_q contains a ball with radius ε. Then by applying Lemma D.1 to each A_q, for $m \le (pn)^d$ 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 $A_0 = \bigcap_{q=1}^Q A_q^{m,1/n}$; then $A \subset A^0$ and $\bigcap_{q=1}^Q A_q^{m,-\varepsilon}$ is empty. By the latter fact, together with condition (M.1$''$) and Lemma A.1, we have $\mathrm{P}(S_n^Y \in A^0) \le C(\log^{1/2}(pn))/n$. Moreover, since S_n^{eX} is Gaussian conditional on X_1^n, by condition (M.1$''$) and Lemma A.1, the inequality $\mathrm{P}(S_n^{eX} \in A^0 \mid X_1^n) \le C(\log^{1/2}(pn))/n$ holds on the event $\Delta_{n,r} \le C_1\overline{\Delta}_n$ since $C_1\overline{\Delta}_n \le b/2$. Since $A \subset A^0$, we conclude that on the event $\Delta_{n,r} \le C_1\overline{\Delta}_n$, $|\mathrm{P}(S_n^{eX} \in A \mid X_1^n) - \mathrm{P}(S_n^Y \in A)| \le C(\log^{1/2}(pn))/n$. This completes the proof since $\mathrm{P}(\Delta_{n,r} > C_1\overline{\Delta}_n) \le \alpha/2$.

Proof of Theorem 4.2. By the triangle inequality, $\rho_n^{\mathrm{EB}}(\mathcal{A}^{\mathrm{re}}) \le \rho_n^{\mathrm{MB}}(\mathcal{A}^{\mathrm{re}}) + \varrho_n^{\mathrm{EB}}(\mathcal{A}^{\mathrm{re}})$, where

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

Also conditional on X_1^n, $X_1^* - \bar{X}, \dots, X_n^* - \bar{X}$ are i.i.d. with mean zero and covariance matrix $\widehat{\Sigma}$. 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 $\widehat{\Sigma}$. 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 $\varrho_n^{\mathrm{EB}}(A^{\mathrm{re}})$ on the event $\{\mathbb{E}_n[(X_{ij} - \bar{X}_j)^2] \ge b$ 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 Proposition 4.3. Here, c, C are constants depending only on b and K under (E.1), and on b, q and K under (E.2); their values may change from place to place. We first note that, for sufficiently small $c > 0$, we may assume that

$$B_n^2 \log^7(pn) \le cn, \tag{41}$$

since otherwise we can make the assertion of the lemma trivial by setting C sufficiently large. To prove the proposition, we will apply Theorem 4.2 separately under (E.1) and under (E.2).

Case (E.1). With (41) in mind, by the proof of Proposition 4.1, we see that $\mathrm{P}(\Delta_{n,r} > b/2) \le \alpha/6$, so that with probability larger than $1 - \alpha/6$, $b/2 \le \mathbb{E}_n[(X_{ij} - \bar{X}_j)^2] \le CB_n$ for all $j = 1, \dots, p$. We turn to bounding $\widehat{L}_n$. Using the inequality $|a - b|^3 \le 4(|a|^3 + |b|^3)$ together with Jensen's inequality, we have

$$\widehat{L}_n \le 4\Bigl(\max_{1\le j\le p} \mathbb{E}_n\bigl[|X_{ij}|^3\bigr] + \max_{1\le j\le p} |\bar{X}_j|^3\Bigr) \le 8 \max_{1\le j\le p} \mathbb{E}_n\bigl[|X_{ij}|^3\bigr].$$

By Lemma E.3,

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

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

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

Taking $t=\log(18/\alpha)\le C\log(pn)$, we conclude that, with $\overline{L}_n=CB_n$ [recall (41)], $\mathrm{P}(\widehat{L}_n>\overline{L}_n)\le\alpha/6$.

Next, consider $\widehat{M}_{n,X}(\phi_n)$. Observe that

$$\max_{1\le j\le p}|X_{ij}-\bar{X}_j|\le 2\max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|,$$

so that

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

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

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

Observe that

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

so that using (41) with c being sufficiently small, we conclude that

$$\mathrm{P}\big(\widehat{M}_{n,X}(\phi_n)>0\big)\le 2(pn)\exp\Big(-\frac{n^{1/3}}{8CB_n^{2/3}\log^{1/3}(pn)}\Big)$$
$$\le 2(pn)\exp\Big(-\frac{n^{1/3}}{8CB_n^{2/3}\log^{7/3}(pn)}\cdot\log^2(pn)\Big)$$
$$\le 2(pn)\exp\Big(-\frac{\log(pn)\log(1/\alpha)}{8c^{1/3}CK}\Big)\le\alpha/6.$$

To bound $\widehat{M}_{n,Y}(\phi_n)$, observe that conditional on $X_1,\dots,X_n$, $\|S_{nj}^{eX}\|_{\psi_2}\le CB_n^{1/2}$ for all $j=1,\dots,p$ on the event $\max_{1\le j\le p}\mathbb{E}_n[(X_{ij}-\bar{X}_j)^2]\le CB_n$, which holds with probability larger than $1-\alpha/6$. Hence, employing the same argument as that used to bound $\widehat{M}_{n,X}(\phi_n)$, we conclude that

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

which implies that

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

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

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

The final conclusion follows from Proposition 4.1.

Case (E.2) In this case, in addition to (41), we may assume that

$$\frac{B_n^2 \log^3(pn)}{\alpha^{2/q} n^{1-2/q}} \le c \le 1 \tag{42}$$

for sufficiently small $c > 0$, since otherwise the assertion of the proposition is trivial by setting C sufficiently large. Then as in the previous case, by the proof of Proposition 4.1, with probability larger than $1-\alpha/6$, $b/2 \le \mathbb{E}_n[(X_{ij}-\bar{X}_j)^2] \le CB_n$ for all $j=1,\dots,p$.

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

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

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

$$\begin{aligned}\mathrm{P}\big(\widehat{L}_n \ge C\big(B_n + B_n^3 n^{-1+3/q}\log p\big) + n^{-1}t\big) &\le Ct^{-q/3}\mathrm{E}\Big[\max_{i,j}|X_{ij}|^q\Big] \\ &\le Ct^{-q/3} n B_n^q.\end{aligned}$$

Solving $Ct^{-q/3}nB_n^q = \alpha/6$, we conclude that $\mathrm{P}(\widehat{L}_n \ge \overline{L}_n) \le \alpha/6$ where $\overline{L}_n = C(B_n + B_n^3 n^{-1+3/q}\alpha^{-3/q}\log p)$.

Next, consider $\widehat{M}_{n,X}(\phi_n)$. As in the previous case,

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

Since the right-hand side is nondecreasing in ϕ_n, and

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

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

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

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

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

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

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

and because of (42), the second term inside the bracket on the right-hand side is at most

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

This completes the proof in this case.

Acknowledgments. We are grateful to Evarist Giné, Friedrich Götze, Ramon van Handel, Vladimir Koltchinskii, Richard Nickl, Larry Wasserman, Galyna Livshyts and Karim Lounici for useful discussions.

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. MR2424985

[2] ADAMCZAK, R. (2010). A few remarks on the operator norm of random Toeplitz matrices. *J. Theoret. Probab.* **23** 85–108. MR2591905

[3] BALL, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. MR1243336

[4] BARVINOK, A. (2014). Thrifty approximations of convex bodies by polytopes. *Int. Math. Res. Not. IMRN* **16** 4341–4356. MR3250035

[5] BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117

[6] BENTKUS, V. YU. (1985). Lower bounds for the rate of convergence in the central limit theorem in Banach spaces. *Lith. Math. J.* **25** 312–320. MR0823198

[7] BENTKUS, V. YU. (1986). Dependence of the Berry–Esseen estimate on the dimension [in Russian]. *Litovsk. Mat. Sb.* **26** 205–210. MR0862741

[8] BHATTACHARYA, R. and RAO, R. (1986). *Normal Approximation and Asymptotic Expansions*. Wiley, New York. MR0855460

[9] BHATTACHARYA, R. N. (1975). On errors of normal approximation. *Ann. Probab.* **3** 815–828. MR0467879

[10] BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386. MR0751577

[11] BORELL, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. MR0388475

[12] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence, with a Foreword by Michel Ledoux*. Oxford Univ. Press, Oxford. MR3185193

[13] CHATTERJEE, S. (2005). A simple invariance theorem. Preprint. Available at arXiv:math/0508213.

[14] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976

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

[16] CHEN, L. and FANG, X. (2011). Multivariate normal approximation by Stein's method: The concentration inequality approach. Preprint. Available at arXiv:1111.4073.

[17] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448

[18] 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.

[19] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

[20] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040

[21] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712

[22] EINMAHL, U. and LI, D. (2008). Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.* **360** 6677–6693. MR2434306

[23] GOLDSTEIN, L. and RINOTT, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949

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

[25] KLIVANS, A., O'DONNELL, R. and SERVEDIO, R. (2008). Learning geometric concepts via Gaussian surface area. In 49*th Annual IEEE Symposium on Foundations of Computer Science*. Philadelphia, PA.

[26] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, Berlin. MR1102015

[27] MASSART, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–884. MR1782276

[28] MILMAN, V. D. and SCHECHTMAN, G. (1986). *Asymptotic Theory of Finite-Dimensional Normed Spaces. Lecture Notes in Math.* **1200**. Springer, Berlin. MR0856576

[29] NAGAEV, S. V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory* (*Tashkent*, 1975). *Lecture Notes in Math.* **550** 419–438. Springer, Berlin. MR0443043

[30] 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 Math.* **1807** 169–187. Springer, Berlin. MR2083397

[31] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model*. Springer, New York. MR3052333

[32] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

[33] REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554

[34] RÖLLIN, A. (2013). Stein's method in high dimensions with applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 529–549. MR3088380

[35] SAZONOV, V. V. (1968). On the multi-dimensional central limit theorem. *Sankhyā Ser. A* **30** 181–204. MR0236979

[36] SAZONOV, V. V. (1981). *Normal Approximation—Some Recent Advances. Lecture Notes in Math.* **879**. Springer, Berlin. MR0643968

[37] SENATOV, V. V. (1980). Several estimates of the rate of convergence in the multidimensional central limit theorem. *Dokl. Akad. Nauk SSSR* **254** 809–812. MR0589638

[38] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** 463–501. MR0133183

[39] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[40] SWEETING, T. J. (1977). Speeds of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28–41. MR0428400

[41] TALAGRAND, M. (2003). *Spin Glasses: A Challenge for Mathematicians*. Springer, Berlin. MR1993891

[42] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

V. CHERNOZHUKOV
DEPARTMENT OF ECONOMICS AND
OPERATIONS RESEARCH CENTER
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02142
USA
E-MAIL: vchern@mit.edu

D. CHETVERIKOV
DEPARTMENT OF ECONOMICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
BUNCHE HALL, 8283
315 PORTOLA PLAZA
LOS ANGELES, CALIFORNIA 90095
USA
E-MAIL: chetverikov@econ.ucla.edu

GRADUATE SCHOOL OF ECONOMICS
UNIVERSITY OF TOKYO
7-3-1 HONGO, BUNKYO-KU
TOKYO 113-0033
JAPAN
E-MAIL: kkato@e.u-tokyo.ac.jp