cemmap

centre for microdata methods and practice

Comparison and anti-concentration bounds for maxima of Gaussian random vectors

Victor Chernozhukov
Denis Chetverikov
Kengo Kato

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

An ESRC Research Centre

COMPARISON AND ANTI-CONCENTRATION BOUNDS FOR MAXIMA OF GAUSSIAN RANDOM VECTORS

VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND KENGO KATO

ABSTRACT. Slepian and Sudakov-Fernique type inequalities, which compare expectations of maxima of Gaussian random vectors under certain restrictions on the covariance matrices, play an important role in probability theory, especially in empirical process and extreme value theories. Here we give explicit comparisons of expectations of smooth functions and distribution functions of maxima of Gaussian random vectors without any restriction on the covariance matrices. We also establish an anti-concentration inequality for the maximum of a Gaussian random vector, which derives a useful upper bound on the Lévy concentration function for the Gaussian maximum. The bound is dimension-free and applies to vectors with arbitrary covariance matrices. This anti-concentration inequality plays a crucial role in establishing bounds on the Kolmogorov distance between maxima of Gaussian random vectors. These results have immediate applications in mathematical statistics. As an example of application, we establish a conditional multiplier central limit theorem for maxima of sums of independent random vectors where the dimension of the vectors is possibly much larger than the sample size.

1. INTRODUCTION

We derive a bound on the difference in expectations of smooth functions of maxima of finite dimensional Gaussian random vectors. We also derive a bound on the Kolmogorov distance between distributions of these maxima. The key property of these bounds is that they depend on the dimension p of Gaussian random vectors only through $\log p$, and on the max norm of the difference between the covariance matrices of the vectors. These results extend and complement the work of [7] that derived an explicit Sudakov-Fernique type bound on the difference of expectations of maxima of Gaussian random vectors. See also [1], Chapter 2. As an application, we establish a conditional multiplier central limit theorem for maxima of sums of independent random vectors where the dimension of the vectors is possibly much larger than the sample size. In all these results, we allow for arbitrary covariance structures between the coordinates in random vectors, which is plausible

especially in applications to high-dimensional statistics. We stress that the derivation of bounds on the Kolmogorov distance is by no means trivial and relies on a new *anti-concentration* inequality for maxima of Gaussian random vectors, which is another main result of this paper (see Comment 4 for what anti-concentration inequalities here precisely refer to and how they differ from the concentration inequalities). These anti-concentration bounds are non-trivial in the following sense: (i) they apply to every dimension p and they are *dimension-free* in the sense that the bounds depend on the dimension p only through the expectation of the maximum of the Gaussian random vector, thereby admitting direct extensions to the infinite dimensional case, namely, separable Gaussian processes (see [10] for this extension and applications to empirical processes). This dimension-free nature is parallel to the Gaussian concentration inequality, which states that the supremum concentrates around the expected supremum. (ii) They allow for arbitrary covariance structures between the coordinates in Gaussian random vectors, and (iii) they are sharp in the sense that there is an example for which the bound is tight up to a dimension independent constant. We note that these anti-concentration bounds are sharper than those that result from application of the universal reverse isoperimetric inequality of [2] (see also [3], p.386-367).

Comparison inequalities for Gaussian random vectors play an important role in probability theory, especially in empirical process and extreme value theories. We refer the reader to [24], [14], [12], [16], [17], [18], [7], and [30] for standard references on this topic. The anti-concentration phenomenon has attracted considerable interest in the context of random matrix theory and the Littlewood-Offord problem in number theory. See, for example, [22], [23], and [29] who remarked that *"concentration is better understood than anti-concentration"*. Those papers were concerned with the anti-concentration in the Euclidean norm for sums of independent random vectors, and the topic and the proof technique here are substantially different from theirs.

Either of the comparison or anti-concentration bounds derived in the paper have many immediate statistical applications, especially in the context of high-dimensional statistical inference, where the dimension p of vectors of interest is much larger than the sample size (see [5] for a textbook treatment of the recent developments of high-dimensional statistics). In particular, the results established here are helpful in deriving an invariance principle for sums of high-dimensional random vectors, and also in establishing the validity of the multiplier bootstrap for inference in practice. We refer the reader to our companion paper [9], where the results established here are applied in several important statistical problems, particularly the analysis of Dantzig selector of [6] in the non-Gaussian setting.

The proof strategy for our anti-concentration inequalities is to directly bound the density function of the maximum of a Gaussian random vector. The paper by [20] is concerned with bounding such a density (see [20], Proposition 3.12) but under positive covariances restriction. This is related

to but different from our anti-concentration bounds. The crucial assumption in their Proposition 3.12 is positivity of all the covariances between the coordinates in the Gaussian random vector, which does not hold in our targeted applications in high-dimensional statistics, for example, analysis of Danzig selector. Moreover, their upper bound on the density depends on the inverse of the lower bound on the covariances – and hence, for example, if there are two independent coordinates in the Gaussian random vector, then the upper bound becomes infinite. Our anti-concentration bounds do not require such positivity (or other) assumptions on covariances and hence are not implied by the results of [20]. Another method for deriving reverse isoperimetric inequalities is to use geometric results of [19], as shown by [13], which leads to dimension-dependent anti-concentration inequalities, which are essentially different from ours. Moreover, our density-bounding proof technique is substantially different from that of [20] based on Malliavin calculus or [19] based on geometric arguments.

The rest of the paper is organized as follows. In Section 2, we present comparison bounds for Gaussian random vectors and its application, namely the conditional multiplier central limit theorem. In Section 3, we present anti-concentration bounds for maxima of Gaussian random vectors. In Sections 4 and 5, we give proofs of the theorems in Sections 2 and 3. The Appendix contains a proof of a technical lemma.

Notation. Denote by $(\Omega, \mathcal{F}, \mathrm{P})$ the underlying probability space. For $a, b \in \mathbb{R}$, we write $a_+ = \max\{0, a\}$ and $a \vee b = \max\{a, b\}$. Let $1(\cdot)$ denote the indicator function. The transpose of a vector z is denoted by z^T. For a function $g : \mathbb{R} \to \mathbb{R}$, we use the notation $\|g\|_\infty = \sup_{z \in \mathbb{R}} |g(z)|$. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and distribution functions of the standard Gaussian distribution, respectively: $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ and $\Phi(x) = \int_{-\infty}^x \phi(t)dt$.

2. Comparison Bounds and Multiplier Bootstrap

2.1. **Comparison bounds.** Let $X = (X_1, \ldots, X_p)^T$ and $Y = (Y_1, \ldots, Y_p)^T$ be centered Gaussian random vectors in $\mathbb{R}^p$ with positive semi-definite covariance matrices $\Sigma^X = (\sigma_{jk}^X)_{1 \le j,k \le p}$ and $\Sigma^Y = (\sigma_{jk}^Y)_{1 \le j,k \le p}$, respectively. The purpose of this section is to give error bounds on the difference of the expectations of smooth functions and the distribution functions of

$$\max_{1 \le j \le p} X_j \quad \text{and} \quad \max_{1 \le j \le p} Y_j$$

in terms of p,

$$\Delta := \max_{1 \le j,k \le p} |\sigma_{jk}^X - \sigma_{jk}^Y|, \text{ and } a_p := \mathrm{E}[\max_{1 \le j \le p} (Y_j/\sigma_{jj}^Y)].$$

The problem of comparing distributions of maxima is of intrinsic difficulty since the maximum function $z = (z_1, \ldots, z_p)^T \mapsto \max_{1 \le j \le p} z_j$ is non-differentiable. To circumvent the problem, we use a smooth approximation

of the maximum function. For $z = (z_1, \ldots, z_p)^T \in \mathbb{R}^p$, consider the function:

$$F_\beta(z) := \beta^{-1} \log \left(\sum_{j=1}^p \exp(\beta z_j) \right),$$

which approximates the maximum function, where $\beta > 0$ is the smoothing parameter that controls the level of approximation (we call this function the "smooth max function"). Indeed, an elementary calculation shows that for every $z \in \mathbb{R}^p$,

$$0 \le F_\beta(z) - \max_{1 \le j \le p} z_j \le \beta^{-1} \log p. \tag{1}$$

This smooth max function arises in the definition of "free energy" in spin glasses. See, for example, [27] and [21]. Here is the first theorem of this section.

Theorem 1 (Comparison bounds for smooth functions). *For every $g \in C^2(\mathbb{R})$ with $\|g'\|_\infty \vee \|g''\|_\infty < \infty$ and every $\beta > 0$,*

$$|\mathrm{E}[g(F_\beta(X)) - g(F_\beta(Y))]| \le (\|g''\|_\infty/2 + \beta\|g'\|_\infty)\Delta,$$

and hence

$$|\mathrm{E}[g(\max_{1 \le j \le p} X_j) - g(\max_{1 \le j \le p} Y_j)]| \le (\|g''\|_\infty/2 + \beta\|g'\|_\infty)\Delta + 2\beta^{-1}\|g'\|_\infty \log p.$$

Proof. See Section 4. $\square$

Comment 1. Minimizing the second bound with respect to $\beta > 0$, we have

$$|\mathrm{E}[g(\max_{1 \le j \le p} X_j) - g(\max_{1 \le j \le p} Y_j)]| \le \|g''\|_\infty \Delta/2 + 2\|g'\|_\infty \sqrt{2\Delta \log p}.$$

This result extends the work of [7], which derived the following Sudakov-Fernique type bound on the expectation of the difference between two Gaussian maxima:

$$|\mathrm{E}[\max_{1 \le j \le p} X_j - \max_{1 \le j \le p} Y_j]| \le 2\sqrt{2\Delta \log p}.$$

Theorem 1 is not applicable to functions of the form $g(z) = 1(z \le x)$ and hence does not directly lead to a bound on the Kolmogorov distance between $\max_{1 \le j \le p} X_j$ and $\max_{1 \le j \le p} Y_j$ (recall that the Kolmogorov distance between (the distributions) of two real valued random variables ξ and η is defined by $\sup_{x \in \mathbb{R}} |\mathrm{P}(\xi \le x) - \mathrm{P}(\eta \le x)|$). Nevertheless, we have the following bounds on the Kolmogorov distance. Recall $a_p = \mathrm{E}[\max_{1 \le j \le p}(Y_j/\sigma_{jj}^Y)]$.

Theorem 2 (Comparison of distributions). *Suppose that $p \ge 2$ and $\sigma_{jj}^Y > 0$ for all $1 \le j \le p$. Then*

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(\max_{1 \le j \le p} X_j \le x) - \mathrm{P}(\max_{1 \le j \le p} Y_j \le x)|$$

$$\le C\Delta^{1/3} \left\{ 1 \vee a_p^2 \vee \log(1/\Delta) \right\}^{1/3} \log^{1/3} p, \tag{2}$$

where $C > 0$ depends only on $\min_{1 \leq j \leq p} \sigma_{jj}^Y$ and $\max_{1 \leq j \leq p} \sigma_{jj}^Y$ (the right side is understood to be 0 when $\Delta = 0$). Moreover, in the worst case, $a_p \leq \sqrt{2 \log p}$, so that

$$\sup_{x \in \mathbb{R}} |P(\max_{1 \leq j \leq p} X_j \leq x) - P(\max_{1 \leq j \leq p} Y_j \leq x)| \leq C' \Delta^{1/3} \{1 \vee \log(p/\Delta)\}^{2/3},$$

where as before $C' > 0$ depends only on $\min_{1 \leq j \leq p} \sigma_{jj}^Y$ and $\max_{1 \leq j \leq p} \sigma_{jj}^Y$.

Proof. See Section 4. □

The first bound (2) is generally sharper than the latter. To see this, consider the simple case where $a_p = O(1)$ as $p \to \infty$, which would happen, for example, when $Y_1, \ldots, Y_p$ come from discretization of a single continuous Gaussian process. Then the right side on (2) is $o(1)$ if $\Delta(\log p) \log \log p = o(1)$, while the second bound requires $\Delta(\log p)^2 = o(1)$.

Comment 2 (On the proof strategy). Bounding the Kolmogorov distance between $\max_{1 \leq j \leq p} X_j$ and $\max_{1 \leq j \leq p} Y_j$ is not immediate from Theorem 1 and this step relies on the anti-concentration inequality for the maximum of a Gaussian random vector, which we will study in Section 3. More formally, by smoothing the indicator and maximum functions, we obtain from Theorem 1 a bound of the following form:

$$\inf_{\beta, \delta > 0} \{\mathcal{L}(\max_{1 \leq j \leq p} Y_j, \beta^{-1} \log p + \delta) + C(\delta^{-2} + \beta \delta^{-1})\Delta\},$$

where $\mathcal{L}(\max_{1 \leq j \leq p} Y_j, \epsilon)$ is the Lévy concentration function for $\max_{1 \leq j \leq p} Y_j$ (see Definition 1 in Section 3 for the formal definition), and $\beta, \delta > 0$ are smoothing parameters (see equations (10) and (11) in the proof of Theorem 2 given in Section 4 for the derivation of the above bound). The bound (2) then follows from bounding the Lévy concentration function by using the anti-concentration inequality derived in Section 3, and optimizing the bound with respect to β, δ.

The proof of Theorem 2 is substantially different from the ("textbook") proof of classical Slepian's inequality. The simplest form of Slepian's inequality states that

$$P(\max_{1 \leq j \leq p} X_j \leq x) \leq P(\max_{1 \leq j \leq p} Y_j \leq x), \ \forall x \in \mathbb{R},$$

whenever $\sigma_{jj}^X = \sigma_{jj}^Y$ and $\sigma_{jk}^X \leq \sigma_{jk}^Y$ for all $1 \leq j, k \leq p$. This inequality is immediately deduced from the following expression:

$$P(\max_{1 \leq j \leq p} X_j \leq x) - P(\max_{1 \leq j \leq p} Y_j \leq x)$$

$$= \sum_{1 \leq j < k \leq p} (\sigma_{jk}^X - \sigma_{jk}^Y) \int_0^1 \left\{ \int_{-\infty}^x \cdots \int_{-\infty}^x \frac{\partial^2 f_t(z)}{\partial z_j \partial z_k} dz \right\} dt, \quad (3)$$

where we assume Σ^X and Σ^Y are non-singular, and $\sigma_{jj}^X = \sigma_{jj}^Y, 1 \leq \forall j \leq p$. Here f_t denotes the density function of $N(0, t\Sigma^X + (1 - t)\Sigma^Y)$. See [14], p.82, for this expression. The expression (3) is of importance and indeed a

source of many interesting probabilistic results (see, for example, [18] and [30] for recent related works). It is not clear (or at least non-trivial), however, whether a bound similar in nature to Theorem 2 can be deduced from the expression (3) when there is no restriction on the covariance matrices except for the condition that $\sigma_{jj}^X = \sigma_{jj}^Y, 1 \leq \forall j \leq p$, and here we take the different route.

The key features of Theorem 2 are: (i) the bound on the Kolmogorov distance between the maxima of Gaussian random vectors in $\mathbb{R}^p$ depends on the dimension p only through $\log p$ and the maximum difference of the covariance matrices Δ, and (ii) it allows for arbitrary covariance matrices for X and Y (except for the nondegeneracy condition that $\sigma_{jj}^Y > 0, \ 1 \leq \forall j \leq p$). These features have an important implication to statistical applications, as discussed below.

2.2. **Conditional multiplier central limit theorem.** Consider the following problem. Suppose that n independent centered random vectors in $\mathbb{R}^p$ of observations $Z_1, \ldots, Z_n$ are given. Here $Z_1, \ldots, Z_n$ are generally non-Gaussian, and the dimension p is allowed to increase with n (that is, the case where $p = p_n \to \infty$ as $n \to \infty$ is allowed). We suppress the possible dependence of p on n for the notational convenience. Suppose that each Z_i has a finite covariance matrix $\mathrm{E}[Z_i Z_i^T]$. Consider the following normalized sum:

$$S_n := (S_{n,1}, \ldots, S_{n,p})^T = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i.$$

The problem here is to approximate the distribution of $\max_{1 \leq j \leq p} S_{n,j}$.

Statistics of this form arise frequently in modern statistical applications. The exact distribution of $\max_{1 \leq j \leq p} S_{n,j}$ is generally unknown. An intuitive idea to approximate the distribution of $\max_{1 \leq j \leq p} S_{n,j}$ is to use the Gaussian approximation. Let $V_1, \ldots, V_n$ be independent Gaussian random vectors in $\mathbb{R}^p$ such that $V_i \sim N(0, \mathrm{E}[Z_i Z_i^T])$, and define

$$T_n := (T_{n,1}, \ldots, T_{n,p}) := \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i \sim N(0, n^{-1} \sum_{i=1}^n \mathrm{E}[Z_i Z_i^T]).$$

It is expected that the distribution of $\max_{1 \leq j \leq p} T_{n,j}$ is close to that of $\max_{1 \leq j \leq p} S_{n,j}$ in the following sense:

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(\max_{1 \leq j \leq p} S_{n,j} \leq x) - \mathrm{P}(\max_{1 \leq j \leq p} T_{n,j} \leq x)| \to 0, \ n \to \infty. \tag{4}$$

When p is fixed, (4) will follow from the classical Lindeberg-Feller central limit theorem, subject to the Lindeberg conditions. The recent paper by [9] established conditions under which this Gaussian approximation (4) holds even when p is comparable or much larger than n. For example, [9] proved that if $c_1 \leq n^{-1} \sum_{i=1}^n \mathrm{E}[Z_{ij}^2] \leq C_1$ and $\mathrm{E}[\exp(|Z_{ij}|/C_1)] \leq 2$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$ for some $0 < c_1 < C_1$, then (4) holds as long as $\log p = o(n^{1/7})$.

The Gaussian approximation (4) is in itself an important step, but in the general case where the covariance matrix $n^{-1} \sum_{i=1}^{n} \mathrm{E}[Z_i Z_i^T]$ is unknown, it is not directly applicable for purposes of statistical inference. In such cases, the following *multiplier bootstrap* procedure will be useful. Let $\eta_1, \ldots, \eta_n$ be independent standard Gaussian random variables independent of $Z_1^n := \{Z_1, \ldots, Z_n\}$. Consider the following randomized sum:

$$S_n^\eta := (S_{n,1}^\eta, \ldots, S_{n,p}^\eta)^T := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i Z_i.$$

Since conditional on Z_1^n,

$$S_n^\eta \sim N(0, n^{-1} \textstyle\sum_{i=1}^{n} Z_i Z_i^T),$$

it is natural to expect that the conditional distribution of $\max_{1 \leq j \leq p} S_{n,j}^\eta$ is "close" to the distribution of $\max_{1 \leq j \leq p} T_{n,j}$ and hence that of $\max_{1 \leq j \leq p} S_{n,j}$. Note here that the conditional distribution of S_n^η is completely known, which makes this distribution useful for purposes of statistical inference. The following proposition makes this intuition rigorous.

Proposition 1 (Conditional multiplier central limit theorem). *Work with the setup as described above. Suppose that $p \geq 2$ and there are some constants $0 < c_1 < C_1$ such that $c_1 \leq n^{-1} \sum_{i=1}^{n} \mathrm{E}[Z_{ij}^2] \leq C_1$ for all $1 \leq j \leq p$. Moreover, suppose that $\widehat{\Delta} := \max_{1 \leq j,k \leq p} |n^{-1} \sum_{i=1}^{n} (Z_{ij} Z_{ik} - \mathrm{E}[Z_{ij} Z_{ik}])|$ obeys the following conditions: as $n \to \infty$,*

$$\widehat{\Delta} (\mathrm{E}[\max_{1 \leq j \leq p} T_{n,j}])^2 \log p = o_\mathrm{P}(1), \quad \widehat{\Delta}(\log p)(1 \vee \log \log p) = o_\mathrm{P}(1). \quad (5)$$

Then we have

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(\max_{1 \leq j \leq p} S_{n,j}^\eta \leq x \mid Z_1^n) - \mathrm{P}(\max_{1 \leq j \leq p} T_{n,j} \leq x)| \xrightarrow{\mathrm{P}} 0, \ \ as \ n \to \infty. \quad (6)$$

Here recall that p is allowed to increase with n.

Proof. Follows immediately from Theorem 2. □

We call this result a "conditional multiplier central limit theorem," where the terminology follows that in empirical process theory. See [28], Chapter 2.9. The notable features of this proposition, which inherit from the features of Theorem 2 discussed above, are: (i) (6) can hold even when p is much larger than n, and (ii) it allows for arbitrary covariance matrices for Z_i (except for the mild scaling condition that $c_1 \leq n^{-1} \sum_{i=1}^{n} \mathrm{E}[Z_{ij}^2] \leq C_1$). The second point is clearly desirable in statistical applications as the information on the true covariance structure is generally (but not always) unavailable. For the first point, we have the following estimate on $\mathrm{E}[\widehat{\Delta}]$.

Lemma 1. *Let $p \geq 2$. There exists a universal constant $C > 0$ such that*

$$\mathrm{E}[\widehat{\Delta}] \leq C \left[\max_{1 \leq j \leq p} (n^{-1} \textstyle\sum_{i=1}^{n} \mathrm{E}[Z_{ij}^4])^{1/2} \sqrt{\frac{\log p}{n}} + (\mathrm{E}[\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} Z_{ij}^4])^{1/2} \frac{\log p}{n} \right].$$

Proof. See the Appendix.   □

Hence with help of Lemma 2.2.2 in [28], we can find various primitive conditions under which (5) holds.

Example 1. Consider the following examples. Here for the sake of simplicity, we use the worst case bound $\mathrm{E}[\max_{1\leq j\leq p} T_{n,j}] \leq \sqrt{2C_1 \log p}$, so that conditions (5) reduce to $\widehat{\Delta} = o_\mathrm{P}((\log p)^{-2})$.

Case (a): Suppose that $\mathrm{E}[\exp(|Z_{ij}|/C_1)] \leq 2$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$ for some $C_1 > 0$. In this case, it is not difficult to verify that $\widehat{\Delta} = o_\mathrm{P}((\log p)^{-2})$ as soon as $\log p = o(n^{1/5})$.

Case (b): Another type of Z_{ij} which arises in regression applications is of the form $Z_{ij} = \varepsilon_i x_{ij}$ where ε_i are stochastic with $\mathrm{E}[\epsilon_i] = 0$ and $\max_{1\leq i\leq n} \mathrm{E}[|\varepsilon_i|^{4q}] = O(1)$ for some $q \geq 1$, and x_{ij} are non-stochastic (typically, ε_i are "errors" and x_{ij} are "regressors"). Suppose that x_{ij} are normalized in such a way that $n^{-1}\sum_{i=1}^n x_{ij}^2 = 1$, and there are bounds $B_n \geq 1$ such that $\max_{1\leq i\leq n}\max_{1\leq j\leq p} |x_{ij}| \leq B_n$, where we allow $B_n \to \infty$. In this case, $\widehat{\Delta} = o_\mathrm{P}((\log p)^{-2})$ as soon as

$$\max\{B_n^2(\log p)^5, B_n^{4q/(2q-1)}(\log p)^{6q/(2q-1)}\} = o(n),$$

since $\max_{1\leq j\leq p}(n^{-1}\sum_{i=1}^n \mathrm{E}[(\varepsilon_i x_{ij})^4]) \leq B_n^2 \max_{1\leq i\leq n} \mathrm{E}[\varepsilon_i^4] = O(B_n^2)$ and $\mathrm{E}[\max_{1\leq i\leq n}\max_{1\leq j\leq p}(\varepsilon_i x_{ij})^4] \leq B_n^4 \mathrm{E}[\max_{1\leq i\leq n} \varepsilon_i^4] = O(n^{1/q} B_n^4)$.

Importantly, in these examples, for (6) to hold, p can increase exponentially in some fractional power of n.

3. ANTI-CONCENTRATION BOUNDS

The following theorem provides bounds on the *Lévy concentration function* of the maximum of a Gaussian random vector in $\mathbb{R}^p$, where the terminology is borrowed from [23].

Definition 1 ([23], Definition 3.1). The *Lévy concentration function* of a real valued random variable ξ is defined for $\epsilon > 0$ as

$$\mathcal{L}(\xi, \epsilon) = \sup_{x\in\mathbb{R}} \mathrm{P}(|\xi - x| \leq \epsilon).$$

Theorem 3 (Anti-concentration). *Let $(X_1, \ldots, X_p)^T$ be a centered Gaussian random vector in $\mathbb{R}^p$ with $\sigma_j^2 := \mathrm{E}[X_j^2] > 0$ for all $1 \leq j \leq p$. Moreover, let $\underline{\sigma} := \min_{1\leq j\leq p} \sigma_j, \overline{\sigma} := \max_{1\leq j\leq p} \sigma_j$, and $a_p := \mathrm{E}[\max_{1\leq j\leq p}(X_j/\sigma_j)]$.*
(i) If the variances are all equal, namely $\underline{\sigma} = \overline{\sigma} = \sigma$, then for every $\epsilon > 0$,

$$\mathcal{L}(\max_{1\leq j\leq p} X_j, \epsilon) \leq 4\epsilon(a_p + 1)/\sigma.$$

(ii) If the variances are not equal, namely $\underline{\sigma} < \overline{\sigma}$, then for every $\epsilon > 0$,

$$\mathcal{L}(\max_{1\leq j\leq p} X_j, \epsilon) \leq C\epsilon\{a_p + \sqrt{1 \vee \log(\underline{\sigma}/\epsilon)}\},$$

where $C > 0$ depends only on $\underline{\sigma}$ and $\overline{\sigma}$.

Since $X_j/\sigma_j \sim N(0,1)$, by a standard calculation, we have $a_p \leq \sqrt{2 \log p}$ in the worst case (see, for example, [27], Proposition 1.1.3), so that the following simpler corollary follows immediately from Theorem 3.

Corollary 1. *Let* $(X_1, \ldots, X_p)^T$ *be a centered Gaussian random vector in* $\mathbb{R}^p$ *with* $\sigma_j^2 := \mathrm{E}[X_j^2] > 0$ *for all* $1 \leq j \leq p$. *Let* $\underline{\sigma} := \min_{1 \leq j \leq p} \sigma_j$ *and* $\overline{\sigma} := \max_{1 \leq j \leq p} \sigma_j$. *Then for every* $\epsilon > 0$,

$$\mathcal{L}(\max_{1 \leq j \leq p} X_j, \epsilon) \leq C\epsilon \sqrt{1 \vee \log(p/\epsilon)},$$

where $C > 0$ *depends only on* $\underline{\sigma}$ *and* $\overline{\sigma}$. *When* σ_j *are all equal,* $\log(p/\epsilon)$ *on the right side can be replaced by* $\log p$.

Comment 3 (Anti-concentration vs. small ball probabilities). The problem of bounding the Lévy concentration function $\mathcal{L}(\max_{1 \leq j \leq p} X_j, \epsilon)$ is qualitatively different from the problem of bounding $\mathrm{P}(\max_{1 \leq j \leq p} |X_j| \leq x)$. For a survey on the latter problem, called the "small ball problem", we refer the reader to [17].

Comment 4 (Concentration vs. anti-concentration). *Concentration inequalities* refer to inequalities bounding $\mathrm{P}(|\xi - x| > \epsilon)$ for a random variable ξ (typically x is the mean or median of ξ). See the monograph [15] for a study of the concentration of measure phenomenon. *Anti-concentration inequalities* in turn refer to reverse inequalities, that is, inequalities bounding $\mathrm{P}(|\xi - x| \leq \epsilon)$. Theorem 3 provides anti-concentration inequalities for $\max_{1 \leq j \leq p} X_j$. [29] remarked that "concentration is better understood than anti-concentration". In the present case, the Gaussian concentration inequality (see [15], Theorem 7.1) states that

$$\mathrm{P}(|\max_{1 \leq j \leq p} X_j - \mathrm{E}[\max_{1 \leq j \leq p} X_j]| \geq r) \leq 2e^{-r^2/(2\overline{\sigma}^2)}, \ r > 0,$$

where the mean can be replace by the median. This inequality is well known and dates back to [4] and [26]. To the best of our knowledge, however, the reverse inequalities in Theorem 3 were not known and are new.

Comment 5 (Anti-concentration for maximum of moduli, $\max_{1 \leq j \leq p} |X_j|$). Versions of Theorem 3 and Corollary 1 continue to hold for $\max_{1 \leq j \leq p} |X_j|$. That is, for example, when σ_j are all equal ($\sigma_j = \sigma$), $\mathcal{L}(\max_{1 \leq j \leq p} |X_j|, \epsilon) \leq 4(a_p' + 1)/\sigma$, where $a_p' := \mathrm{E}[\max_{1 \leq j \leq p} |X_j|/\sigma]$. To see this, observe that $\max_{1 \leq j \leq p} |X_j| = \max_{1 \leq j \leq 2p} X_j'$ where $X_j' = X_j$ for $j = 1, \ldots, p$ and $X_{p+j}' = -X_j$ for $j = 1, \ldots, p$. Hence we may apply Theorem 3 to $(X_1', \ldots, X_{2p}')^T$ to obtain the desired conclusion.

Comment 6 (A sketch of the proof of Theorem 3). For the reader's convenience, we provide a sketch of the proof of Theorem 3. We focus here on the simple case where all the variances are equal to one ($\sigma_1 = \cdots = \sigma_p = 1$). Then the distribution of $Z = \max_{1 \leq j \leq p} X_j$ is absolutely continuous and its density can be written as $\phi(z)G(z)$ where the map $z \mapsto G(z)$ is non-decreasing. Consequently, it is then not difficult to see that $G(z) \leq \mathrm{P}(Z >$

$z)/\{1-\Phi(z)\} \leq 2(z \vee 1)e^{-(z-a_p)_+^2/2}/\phi(z)$, where the second inequality follows from Mill's inequality combined with the Gaussian concentration inequality. Hence the density of Z is bounded by $2(z \vee 1)e^{-(z-a_p)_+^2/2}$, which immediately leads to the bound $\mathcal{L}(\max_{1 \leq j \leq p} X_j, \epsilon) \leq 4(a_p + 1)\epsilon$.

In a trivial example where $p = 1$, it is immediate to see that $\mathrm{P}(|X_1 - x| \leq \epsilon) \leq \epsilon\sqrt{2/(\pi\sigma_1^2)}$. A non-trivial case is the situation where $p \to \infty$. In such a case, it is typically not known whether $\max_{1 \leq j \leq p} X_j$ has a limiting distribution as $p \to \infty$, even after normalization (recall that except for $\underline{\sigma} > 0$, we allow for general covariance structures between $X_1, \ldots, X_p$), and therefore it is not trivial at all whether, for every sequence $\epsilon = \epsilon_p \to 0$ (or at some rate), $\mathcal{L}(\max_{1 \leq j \leq p} X_j, \epsilon) \to 0$ or how fast $\epsilon = \epsilon_p \to 0$ should be to guarantee that $\mathcal{L}(\max_{1 \leq j \leq p} X_j, \epsilon) \to 0$. Theorem 3 answers this question with explicit, non-asymptotic bounds.

Importantly, the bounds in Theorem 3 are *dimension-free* in the sense that, similar to the Gaussian concentration inequality, they depend on the dimension p only through a_p – the expectation of the maximum of the (normalized) Gaussian random vector. Hence Theorem 3 admits direct extensions to the infinite dimensional case, namely separable Gaussian processes, as long as the corresponding expectation is finite. See our companion paper [10] for formal treatments and applications of this extension.

The presence of a_p on the bounds is essential and can not be removed in general, as the following example suggests. This shows that there does not exist a substantially sharper estimate of the universal bound of the concentration function than that given in Theorem 3. Potentially, there could be refinements but they would have to rely on the particular (hence non-universal) features of the covariance structure between $X_1, \ldots, X_p$.

Example 2 (Partial converse of Theorem 3). Let $X_1, \ldots, X_p$ be independent standard Gaussian random variables. By Theorem 1.5.3 of [14], as $p \to \infty$,

$$b_p(\max_{1 \leq j \leq p} X_j - d_p) \xrightarrow{d} G(0, 1), \tag{7}$$

where

$$b_p := \sqrt{2 \log p}, \quad d_p := b_p - \frac{\log(4\pi) + \log \log p}{2b_p},$$

and $G(0, 1)$ denotes the standard Gumbel distribution, that is, the distribution having the density $g(x) = e^{-x}e^{-e^{-x}}$ for $x \in \mathbb{R}$. In fact, we can show that the density of $b_p(\max_{1 \leq j \leq p} X_j - d_p)$ converges to that of $G(0, 1)$ locally uniformly. To see this, we begin with noting that the density of $b_p(\max_{1 \leq j \leq p} X_j - d_p)$ is given by

$$g_p(x) = \frac{p}{b_p}\phi(d_p + b_p^{-1}x)[\Phi(d_p + b_p^{-1}x)]^{p-1}.$$

Pick any $x \in \mathbb{R}$. Since, by the weak convergence result (7),

$$[\Phi(d_p + b_p^{-1}x)]^p = \mathrm{P}(b_p(\max_{1 \leq j \leq p} X_j - d_p) \leq x) \to e^{-e^{-x}}, \ p \to \infty,$$

we have $[\Phi(d_p + b_p^{-1}x)]^{p-1} \to e^{-e^{-x}}$. Hence it remains to show that

$$\frac{p}{b_p}\phi(d_p + b_p^{-1}x) \to e^{-x}.$$

Taking the logarithm of the left side yields

$$\log p - \log b_p - \log(\sqrt{2\pi}) - (d_p + b_p^{-1}x)^2/2. \tag{8}$$

Expanding $(d_p + b_p^{-1}x)^2$ gives that

$$d_p^2 + 2d_p b_p^{-1}x + b_p^{-2}x^2 = b_p^2 - \log\log p - \log(4\pi) + 2x + o(1), \ p \to \infty,$$

by which we have (8) $= -x + o(1)$. This shows that $g_p(x) \to g(x)$ for all $x \in \mathbb{R}$. Moreover, this convergence takes place locally uniformly in x, that is, for every $K > 0$, $g_p(x) \to g(x)$ uniformly in $x \in [-K, K]$.

On the other hand, the density of $\max_{1 \le j \le p} X_j$ is given by $f_p(x) = p\phi(x)[\Phi(x)]^{p-1}$. By this form, for every $K > 0$, there exist a constant $c > 0$ and a positive integer p_0 depending only on K such that for $p \ge p_0$,

$$\inf_{x \in [d_p - Kb_p^{-1}, d_p + Kb_p^{-1}]} b_p^{-1}f_p(x) = \inf_{x \in [-K,K]} g_p(x) \ge \inf_{x \in [-K,K]} g(x) + o(1) \ge c,$$

which shows that for $p \ge p_0$,

$$f_p(x) \ge cb_p, \ \forall x \in [d_p - Kb_p^{-1}, d_p + Kb_p^{-1}].$$

Therefore, we conclude that for $p \ge p_0$,

$$\mathrm{P}(|\max_{1 \le j \le p} X_j - d_p| \le \epsilon) = \int_{d_p - \epsilon}^{d_p + \epsilon} f_p(x)dx \ge 2c\epsilon b_p, \ \forall \epsilon \in [0, Kb_p^{-1}].$$

By the Gaussian maximal inequality and Lemma 2.3.15 of [11], we have

$$\sqrt{\log p}/12 \le \mathbb{E}[\max_{1 \le j \le p} X_j] \le \sqrt{2\log p}.$$

Hence, by the previous result, for every $K' > 0$, there exist a constant $c' > 0$ and a positive integer p_0' depending only on K' such that for $p \ge p_0'$, $a_p \ge 1$ and

$$\mathcal{L}(\max_{1 \le j \le p} X_j, \epsilon) \ge \mathrm{P}(|\max_{1 \le j \le p} X_j - d_p| \le \epsilon) \ge c'\epsilon a_p, \ \forall \epsilon \in [0, K'a_p^{-1}].$$

$\square$

4. Proofs for Section 2

4.1. **Proof of Theorem 1.** Here for a smooth function $f : \mathbb{R}^p \to \mathbb{R}$, we write $\partial_j f(z) = \partial f(z)/\partial z_j$ for $z = (z_1, \ldots, z_p)^T$. We shall use the following version of Stein's identity.

Lemma 2 (Stein's identity). *Let $W = (W_1, \ldots, W_p)^T$ be a centered Gaussian random vector in $\mathbb{R}^p$. Let $f : \mathbb{R}^p \to \mathbb{R}$ be a C^1-function such that $\mathrm{E}[|\partial_j f(W)|] < \infty$ for all $1 \le j \le p$. Then for every $1 \le j \le p$,*

$$\mathrm{E}[W_j f(W)] = \sum_{k=1}^{p} \mathrm{E}[W_j W_k] \mathrm{E}[\partial_k f(W)].$$

Proof of Lemma 2. See Section A.6 of [27]; also [8] and [25]. □

We will use the following properties of the smooth max function.

Lemma 3. *For every $1 \le j, k \le p$,*

$$\partial_j F_\beta(z) = \pi_j(z), \quad \partial_j \partial_k F_\beta(z) = \beta w_{jk}(z),$$

where

$$\pi_j(z) := e^{\beta z_j} / \sum_{m=1}^{p} e^{\beta z_m}, \ w_{jk}(z) := 1(j = k)\pi_j(z) - \pi_j(z)\pi_k(z).$$

Moreover,

$$\pi_j(z) \ge 0, \ \sum_{j=1}^{p} \pi_j(z) = 1, \ \sum_{j,k=1}^{p} |w_{jk}(z)| \le 2.$$

Proof of Lemma 3. The first property was noted in [7]. The other properties follow from a direct calculation. □

Lemma 4. *Let $m := g \circ F_\beta$ with $g \in C^2(\mathbb{R})$. Then for every $1 \le j, k \le p$,*

$$\partial_j \partial_k m(z) = (g'' \circ F_\beta)(z)\pi_j(z)\pi_k(z) + \beta(g' \circ F_\beta)(z)w_{jk}(z),$$

where π_j and w_{jk} are defined in Lemma 3.

Proof of lemma 4. The proof follows from a direct calculation. □

Proof of Theorem 1. Without loss of generality, we may assume that X and Y are independent, so that $\mathrm{E}[X_j Y_k] = 0$ for all $1 \le j, k \le p$. Consider the following Slepian interpolation between X and Y:

$$Z(t) := \sqrt{t}X + \sqrt{1-t}Y, \ t \in [0, 1].$$

Let $m := g \circ F_\beta$ and $\Psi(t) := \mathrm{E}[m(Z(t))]$. Then

$$|\mathrm{E}[m(X)] - \mathrm{E}[m(Y)]| = |\Psi(1) - \Psi(0)| = \left| \int_0^1 \Psi'(t)dt \right|.$$

Here we have

$$\Psi'(t) = \frac{1}{2} \sum_{j=1}^{p} \mathrm{E}[\partial_j m(Z(t))(t^{-1/2}X_j - (1-t)^{-1/2}Y_j)]$$

$$= \frac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} (\sigma_{jk}^X - \sigma_{jk}^Y)\mathrm{E}[\partial_j \partial_k m(Z(t))],$$

where the second equality follows from applying Lemma 2 to $W = (t^{-1/2}X_j - (1-t)^{-1/2}Y_j, Z(t)^T)^T$ and $f(W) = \partial_j m(Z(t))$. Hence

$$
\left| \int_0^1 \Psi'(t)dt \right| \le \frac{1}{2} \sum_{j,k=1}^p |\sigma_{jk}^X - \sigma_{jk}^Y| \left| \int_0^1 \mathrm{E}[\partial_j \partial_k m(Z(t))]dt \right|
$$

$$
\le \frac{1}{2} \max_{1 \le j,k \le p} |\sigma_{jk}^X - \sigma_{jk}^Y| \int_0^1 \sum_{j,k=1}^p |\mathrm{E}[\partial_j \partial_k m(Z(t))]| \, dt
$$

$$
= \frac{\Delta}{2} \int_0^1 \sum_{j,k=1}^p |\mathrm{E}[\partial_j \partial_k m(Z(t))]| \, dt.
$$

By Lemmas 3 and 4,

$$
\sum_{j,k=1}^p |\partial_j \partial_k m(Z(t))| \le |(g'' \circ F_\beta)(Z(t))| + 2\beta|(g' \circ F_\beta)(Z(t))|.
$$

Therefore, we have

$$
|\mathrm{E}[g(F_\beta(X)) - g(F_\beta(Y))]|
$$
$$
\le \Delta \times \left\{ \frac{1}{2} \int_0^1 \mathrm{E}[|(g'' \circ F_\beta)(Z(t))|]dt + \beta \int_0^1 \mathrm{E}[|(g' \circ F_\beta)(Z(t))|]dt \right\}
$$
$$
\le \Delta(\|g''\|_\infty/2 + \beta\|g'\|_\infty),
$$

which leads to the first assertion. The second assertion follows from the inequality (1). This completes the proof. $\square$

4.2. **Proof of Theorem 2.** The final assertion follows from the inequality $a_p \le \sqrt{2\log p}$ (see, for example, [27], Proposition 1.1.3). Hence we prove (2). We first note that we may assume that $0 < \Delta < 1$ since otherwise the proof is trivial (take $C \ge 2$ in (2)). In what follows, let $C > 0$ be a generic constant that depends only on $\min_{1 \le j \le p} \sigma_{jj}^Y$ and $\max_{1 \le j \le p} \sigma_{jj}^Y$, and its value may change from place to place. For $\beta > 0$, define $e_{p,\beta} := \beta^{-1}\log p$. Consider and fix a C^2-function $g_0 : \mathbb{R} \to [0,1]$ such that $g_0(t) = 1$ for $t \le 0$ and $g_0(t) = 0$ for $t \ge 1$. For example, we may take

$$
g_0(t) = \begin{cases} 0, & t \ge 1, \\ 30 \int_t^1 s^2(1-s)^2 ds, & 0 < t < 1, \\ 1, & t \le 0. \end{cases}
$$

For given $x \in \mathbb{R}, \beta > 0$ and $\delta > 0$, define $g_{x,\beta,\delta}(t) = g_0(\delta^{-1}(t - x - e_{p,\beta}))$. For this function $g_{x,\beta,\delta}$, $\|g'_{x,\beta,\delta}\|_\infty = \delta^{-1}\|g'_0\|_\infty$ and $\|g''_{x,\beta,\delta}\|_\infty = \delta^{-2}\|g''_0\|_\infty$. Moreover,

$$
1(t \le x + e_{p,\beta}) \le g_{x,\beta,\delta}(t) \le 1(t \le x + e_{p,\beta} + \delta), \ \forall t \in \mathbb{R}. \tag{9}
$$

For arbitrary $x \in \mathbb{R}, \beta > 0$ and $\delta > 0$, observe that

$$
\begin{aligned}
\mathrm{P}(\max_{1 \le j \le p} X_j \le x) &\le \mathrm{P}(F_\beta(X) \le x + e_{p,\beta}) \le \mathrm{E}[g_{x,\beta,\delta}(F_\beta(X))] \\
&\le \mathrm{E}[g_{x,\beta,\delta}(F_\beta(Y))] + C(\delta^{-2} + \beta\delta^{-1})\Delta \\
&\le \mathrm{P}(F_\beta(Y) \le x + e_{p,\beta} + \delta) + C(\delta^{-2} + \beta\delta^{-1})\Delta \\
&\le \mathrm{P}(\max_{1 \le j \le p} Y_j \le x + e_{p,\beta} + \delta) + C(\delta^{-2} + \beta\delta^{-1})\Delta, \quad (10)
\end{aligned}
$$

where the first inequality follows from the inequality (1), the second from the inequality (9), the third from Theorem 1, the fourth from the inequality (9), and the last from the inequality (1). We wish to compare $\mathrm{P}(\max_{1 \le j \le p} Y_j \le x + e_{p,\beta} + \delta)$ with $\mathrm{P}(\max_{1 \le j \le p} Y_j \le x)$, and this is where the anti-concentration inequality plays its role. By Theorem 3, we have

$$
\begin{aligned}
&\mathrm{P}(\max_{1 \le j \le p} Y_j \le x + e_{p,\beta} + \delta) - \mathrm{P}(\max_{1 \le j \le p} Y_j \le x) \\
&= \mathrm{P}(x < \max_{1 \le j \le p} Y_j \le x + e_{p,\beta} + \delta) \le \mathcal{L}(\max_{1 \le j \le p} Y_j, e_{p,\beta} + \delta) \quad (11) \\
&\le C(e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log\{1/(e_{p,\beta} + \delta)\}} \le C(e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log(1/\delta)}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathrm{P}(\max_{1 \le j \le p} X_j \le x) - \mathrm{P}(\max_{1 \le j \le p} Y_j \le x) \\
&\qquad\qquad \le C\left\{(\delta^{-2} + \beta\delta^{-1})\Delta + (e_{p,\beta} + \delta)\sqrt{1 \vee a_p^2 \vee \log(1/\delta)}\right\}. \quad (12)
\end{aligned}
$$

Take $a = a_p \vee \log^{1/2}(1/\Delta)$, and choose β and δ in such a way that

$$
\beta = \delta^{-1}\log p \quad \text{and} \quad \delta = \Delta^{1/3}(1 \vee a)^{-1/3}(2\log p)^{1/3}.
$$

Recall that $p \ge 2$ and $0 < \Delta < 1$. Observe that $(\delta^{-2} + \beta\delta^{-1})\Delta \le C\Delta^{1/3}(1 \vee a)^{2/3}\log^{1/3} p, (e_{p,\beta} + \delta)(1 \vee a_p) \le C\Delta^{1/3}(1 \vee a)^{2/3}\log^{1/3} p$, and since $\delta \ge \Delta^{1/3}(1 \vee a)^{-1/3}$, we have $\log(1/\delta) \le (1/3)\log((1 \vee a)/\Delta)$. Hence the right side on (12) is bounded by $C\Delta^{1/3}\{(1 \vee a)^{2/3}\log^{1/3} p + (1 \vee a)^{-1/3}(\log^{1/3} p)\log^{1/2}((1 \vee a)/\Delta)\}$. In addition, $(1 \vee a)^{-1}\log^{1/2}(1 \vee a)$ is bounded by a universal constant, so that

right side of (12)

$$
\le C\Delta^{1/3}\{(1 \vee a)^{2/3}\log^{1/3} p + (1 \vee a)^{-1/3}(\log^{1/3} p)\log^{1/2}(1/\Delta)\}.
$$

The second term inside the bracket is bounded by $(\log^{1/3} p)\log^{1/3}(1/\Delta)$ as $(1 \vee a)^{-1/3} \le (\log(1/\Delta))^{-1/6}$, and first term is bound by $(1 \vee a_p)^{2/3}\log^{1/3} p + (\log^{1/3} p)\log^{1/3}(1/\Delta)$. Adjusting the constant C, the right side on the above displayed equation is bounded by $C\Delta^{1/3}\{1 \vee a_p^2 \vee \log(1/\Delta)\}^{1/3}\log^{1/3} p$.

For the opposite direction, observe that

$$P(\max_{1 \le j \le p} X_j \le x) \ge P(F_\beta(X) \le x) \ge E[g_{x-e_{p,\beta}-\delta,\beta,\delta}(F_\beta(X))]$$

$$\ge E[g_{x-e_{p,\beta}-\delta,\beta,\delta}(F_\beta(Y))] - C(\delta^{-2} + \beta\delta^{-1})\Delta$$
$$\ge P(F_\beta(Y) \le x - \delta) - C(\delta^{-2} + \beta\delta^{-1})\Delta$$
$$\ge P(\max_{1 \le j \le p} Y_j \le x - e_{p,\beta} - \delta) - C(\delta^{-2} + \beta\delta^{-1})\Delta.$$

The rest of the proof is similar and hence omitted. $\qquad\square$

5. Proof of Theorem 3

The proof of Theorem 3 uses some properties of Gaussian measures. We begin with preparing technical tools. The following two facts were essentially noted in [31, 32] (note: [31] and [32] did not contain a proof of Lemma 5, which we find non-trivial). For the sake of completeness, we give their proofs after the proof of Theorem 3.

Lemma 5. *Let* $(W_1, \ldots, W_p)^T$ *be a (not necessarily centered) Gaussian random vector in* $\mathbb{R}^p$ *with* $\mathrm{Var}(W_j) = 1$ *for all* $1 \le j \le p$. *Suppose that* $\mathrm{Corr}(W_j, W_k) < 1$ *whenever* $j \ne k$. *Then the distribution of* $\max_{1 \le j \le p} W_j$ *is absolutely continuous with respect to the Lebesgue measure and a version of the density is given by*

$$f(x) = \phi(x) \sum_{j=1}^p e^{E[W_j]x - (E[W_j])^2/2} \cdot P\left(W_k \le x, \forall k \ne j \mid W_j = x\right). \qquad (13)$$

Lemma 6. *Let* $(W_0, W_1, \ldots, W_p)^T$ *be a (not necessarily centered) Gaussian random vector with* $\mathrm{Var}(W_j) = 1$ *for all* $0 \le j \le p$. *Suppose that* $E[W_0] \ge 0$. *Then the map*

$$x \mapsto e^{E[W_0]x - (E[W_0])^2/2} \cdot P(W_j \le x, 1 \le \forall j \le p \mid W_0 = x) \qquad (14)$$

is non-decreasing on $\mathbb{R}$.

Let us also recall (a version of) the Gaussian concentration (more precisely, deviation) inequality. See, for example, [15], Theorem 7.1, for its proof.

Lemma 7. *Let* $(X_1, \ldots, X_p)^T$ *be a centered Gaussian random vector in* $\mathbb{R}^p$ *with* $\max_{1 \le j \le p} E[X_j^2] \le \sigma^2$ *for some* $\sigma^2 > 0$. *Then for every* $r > 0$,

$$P(\max_{1 \le j \le p} X_j \ge E[\max_{1 \le j \le p} X_j] + r) \le e^{-r^2/(2\sigma^2)}.$$

We are now in position to prove Theorem 3.

Proof of Theorem 3. The proof consists of three steps.

Step 1. This step reduces the analysis to the unit variance case. Pick any $x \geq 0$. Let $W_j := (X_j - x)/\sigma_j + x/\underline{\sigma}$. Then $\mathrm{E}[W_j] \geq 0$ and $\mathrm{Var}(W_j) = 1$. Define $Z := \max_{1 \leq j \leq p} W_j$. Then we have

$$\mathrm{P}(|\max_{1 \leq j \leq p} X_j - x| \leq \epsilon) \leq \mathrm{P}\left(\left|\max_{1 \leq j \leq p} \frac{X_j - x}{\sigma_j}\right| \leq \frac{\epsilon}{\underline{\sigma}}\right)$$

$$\leq \sup_{y \in \mathbb{R}} \mathrm{P}\left(\left|\max_{1 \leq j \leq p} \frac{X_j - x}{\sigma_j} + \frac{x}{\underline{\sigma}} - y\right| \leq \frac{\epsilon}{\underline{\sigma}}\right) = \sup_{y \in \mathbb{R}} \mathrm{P}\left(|Z - y| \leq \frac{\epsilon}{\underline{\sigma}}\right).$$

Step 2. This step bounds the density of Z. Without loss of generality, we may assume that $\mathrm{Corr}(W_j, W_k) < 1$ whenever $j \neq k$. Since the marginal distribution of W_j is $N(\mu_j, 1)$ where $\mu_j := \mathrm{E}[W_j] = (x/\underline{\sigma} - x/\sigma_j) \geq 0$, by Lemma 5, Z has density of the form

$$f_p(z) = \phi(z)G_p(z), \tag{15}$$

where the map $z \mapsto G_p(z)$ is non-decreasing by Lemma 6. Define $\bar{z} := (1/\underline{\sigma} - 1/\overline{\sigma})x$, so that $\mu_j \leq \bar{z}$ for every $1 \leq j \leq p$. Moreover, define $\bar{Z} := \max_{1 \leq j \leq p}(W_j - \mu_j)$. Then

$$\int_z^\infty \phi(u)du G_p(z) \leq \int_z^\infty \phi(u)G_p(u)du = \mathrm{P}(Z > z)$$

$$\leq \mathrm{P}(\bar{Z} > z - \bar{z}) \leq \exp\left\{-\frac{(z - \bar{z} - \mathrm{E}[\bar{Z}])_+^2}{2}\right\},$$

where the last inequality is due to the Gaussian concentration inequality (Lemma 7). Note that $W_j - \mu_j = X_j/\sigma_j$, so that

$$\mathrm{E}[\bar{Z}] = \mathrm{E}[\max_{1 \leq j \leq p}(X_j/\sigma_j)] =: a_p.$$

Therefore, for every $z \in \mathbb{R}$,

$$G_p(z) \leq \frac{1}{1 - \Phi(z)} \exp\left\{-\frac{(z - \bar{z} - a_p)_+^2}{2}\right\}. \tag{16}$$

Mill's inequality states that for $z > 0$,

$$z \leq \frac{\phi(z)}{1 - \Phi(z)} \leq z\frac{1 + z^2}{z^2},$$

and in particular $(1 + z^2)/z^2 \leq 2$ when $z > 1$. Moreover, $\phi(z)/\{1 - \Phi(z)\} \leq 1.53 \leq 2$ for $z \in (-\infty, 1)$. Therefore,

$$\phi(z)/\{1 - \Phi(z)\} \leq 2(z \vee 1), \ \forall z \in \mathbb{R}.$$

Hence we conclude from this, (16), and (15) that

$$f_p(z) \leq 2(z \vee 1) \exp\left\{-\frac{(z - \bar{z} - a_p)_+^2}{2}\right\}, \ \forall z \in \mathbb{R}.$$

Step 3. By Step 2, for every $y \in \mathbb{R}$ and $t > 0$, we have

$$\mathrm{P}(|Z - y| \leq t) = \int_{y-t}^{y+t} f_p(z)dz \leq 2t \max_{z \in [y-t, y+t]} f_p(z) \leq 4t(\bar{z} + a_p + 1),$$

where the last inequality follows from the fact that the map $z \mapsto ze^{-(z-a)^2/2}$ (with $a > 0$) is non-increasing on $[a+1, \infty)$. Combining this inequality with Step 1, for every $x \geq 0$ and $\epsilon > 0$, we have

$$P(|\max_{1 \leq j \leq p} X_j - x| \leq \epsilon) \leq 4\epsilon\{(1/\underline{\sigma} - 1/\overline{\sigma})|x| + a_p + 1\}/\underline{\sigma}. \qquad (17)$$

This inequality also holds for $x < 0$ by the similar argument, and hence it holds for every $x \in \mathbb{R}$.

If $\underline{\sigma} = \overline{\sigma} = \sigma$, then we have

$$P(|\max_{1 \leq j \leq p} X_j - x| \leq \epsilon) \leq 4\epsilon(a_p + 1)/\sigma, \ \forall x \in \mathbb{R}, \ \forall \epsilon > 0,$$

which leads to the first assertion of the theorem.

On the other hand, consider the case where $\underline{\sigma} < \overline{\sigma}$. Suppose first that $0 < \epsilon \leq \underline{\sigma}$. By the Gaussian concentration inequality (Lemma 7), for $|x| \geq \epsilon + \overline{\sigma}(a_p + \sqrt{2\log(\underline{\sigma}/\epsilon)})$, we have

$$P(|\max_{1 \leq j \leq p} X_j - x| \leq \epsilon) \leq P(\max_{1 \leq j \leq p} X_j \geq |x| - \epsilon)$$

$$\leq P\left(\max_{1 \leq j \leq p} X_j \geq E[\max_{1 \leq j \leq p} X_j] + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\epsilon)}\right) \leq \epsilon/\underline{\sigma}. \qquad (18)$$

For $|x| \leq \epsilon + \overline{\sigma}(a_p + \sqrt{2\log(\underline{\sigma}/\epsilon)})$, by (17) and using $\epsilon \leq \underline{\sigma}$, we have

$$P(|\max_{1 \leq j \leq p} X_j - x| \leq \epsilon)$$

$$\leq 4\epsilon\{(\overline{\sigma}/\underline{\sigma})a_p + (\overline{\sigma}/\underline{\sigma} - 1)\sqrt{2\log(\underline{\sigma}/\epsilon)} + 2 - \underline{\sigma}/\overline{\sigma}\}/\underline{\sigma}. \qquad (19)$$

Combining (18) and (19), we obtain the inequality in (ii) for $0 < \epsilon \leq \underline{\sigma}$ (with a suitable choice of C). If $\epsilon > \underline{\sigma}$, the inequality in (ii) trivially follows by taking $C \geq 1/\underline{\sigma}$. This completes the proof. $\qquad \square$

Proof of Lemma 5. Let $M := \max_{1 \leq j \leq p} W_j$. The absolute continuity of the distribution of M is deduced from the fact that $P(M \in A) \leq \sum_{j=1}^{p} P(W_j \in A)$ for every Borel measurable subset A of $\mathbb{R}$. Hence, to show that a version of the density of M is given by (13), it is enough to show that $\lim_{\epsilon \downarrow 0} \epsilon^{-1} P(x < M \leq x + \epsilon)$ equals the right side on (13) for a.e. $x \in \mathbb{R}$.

For every $x \in \mathbb{R}$ and $\epsilon > 0$, observe that

$$\{x < M \leq x + \epsilon\}$$
$$= \{\exists i_0, W_{i_0} > x \text{ and } \forall i, W_i \leq x + \epsilon\}$$
$$= \{\exists i_1, x < W_{i_1} \leq x + \epsilon \text{ and } \forall i \neq i_1, W_i \leq x\}$$
$$\quad \cup \{\exists i_1, \exists i_2, x < W_{i_1} \leq x + \epsilon, x < W_{i_2} \leq x + \epsilon \text{ and } \forall i \notin \{i_1, i_2\}, W_i \leq x\}$$
$$\qquad \vdots$$
$$\quad \cup \{\forall i, x < W_i \leq x + \epsilon\}$$
$$=: A_1^{x,\epsilon} \cup A_2^{x,\epsilon} \cup \cdots \cup A_p^{x,\epsilon}.$$

Note that the events $A_1^{x,\epsilon}, A_2^{x,\epsilon}, \ldots, A_p^{x,\epsilon}$ are disjoint. For $A_1^{x,\epsilon}$, since
$$A_1^{x,\epsilon} = \cup_{i=1}^p \{x < W_i \leq x + \epsilon \text{ and } W_j \leq x, \forall j \neq i\},$$
where the events on the right side are disjoint, we have
$$\mathrm{P}(A_1^{x,\epsilon}) = \sum_{i=1}^p \mathrm{P}(x < W_i \leq x + \epsilon \text{ and } W_j \leq x, \forall j \neq i)$$
$$= \sum_{i=1}^p \int_x^{x+\epsilon} \mathrm{P}(W_j \leq x, \forall j \neq i \mid W_i = u)\phi(u - \mu_i)du,$$
where $\mu_i := \mathrm{E}[W_i]$. We show that for every $1 \leq i \leq p$ and a.e. $x \in \mathbb{R}$, the map $u \mapsto \mathrm{P}(W_j \leq x, \forall j \neq i \mid W_i = u)$ is right continuous at x. Let $X_j = W_j - \mu_j$ so that X_j are standard Gaussian random variables. Then
$$\mathrm{P}(W_j \leq x, \forall j \neq i \mid W_i = u) = \mathrm{P}(X_j \leq x - \mu_j, \forall j \neq i \mid X_i = u - \mu_i).$$
Pick $i = 1$. Let $V_j = X_j - \mathrm{E}[X_j X_1]X_1$ be the residual from the orthogonal projection of X_j on X_1. Note that the vector $(V_j)_{2 \leq j \leq p}$ and X_1 are jointly Gaussian and uncorrelated, and hence independent, by which we have
$$\mathrm{P}(X_j \leq x - \mu_j, 2 \leq \forall j \leq p \mid X_1 = u - \mu_1)$$
$$= \mathrm{P}(V_j \leq x - \mu_j - \mathrm{E}[X_j X_1](u - \mu_1), 2 \leq \forall j \leq p \mid X_1 = u - \mu_1)$$
$$= \mathrm{P}(V_j \leq x - \mu_j - \mathrm{E}[X_j X_1](u - \mu_1), 2 \leq \forall j \leq p).$$
Define $J := \{j \in \{2, \ldots, p\} : \mathrm{E}[X_j X_1] \leq 0\}$ and $J^c := \{2, \ldots, p\}\backslash J$. Then
$$\mathrm{P}(V_j \leq x - \mu_j - \mathrm{E}[X_j X_1](u - \mu_1), 2 \leq \forall j \leq p)$$
$$\to \mathrm{P}(V_j \leq x_j, \forall j \in J, V_{j'} < x_{j'}, \forall j' \in J^c), \text{ as } u \downarrow x,$$
where $x_j = x - \mu_j - \mathrm{E}[X_j X_1](x - \mu_1)$. Here each V_j either degenerates to 0 (which occurs only when X_j and X_1 are perfectly negatively correlated, that is, $\mathrm{E}[X_j X_1] = -1$) or has a non-degenerate Gaussian distribution, and hence for every $x \in \mathbb{R}$ except for at most $(p-1)$ points $(\mu_1 + \mu_j)/2, 2 \leq j \leq p$,
$$\mathrm{P}(V_j \leq x_j, \forall j \in J, V_{j'} < x_{j'}, \forall j' \in J^c) = \mathrm{P}(V_j \leq x_j, 2 \leq \forall j \leq p)$$
$$= \mathrm{P}(W_j \leq x, 2 \leq \forall j \leq p \mid W_1 = x).$$
Hence for $i = 1$ and a.e. $x \in \mathbb{R}$, the map $u \mapsto \mathrm{P}(W_i \leq x, \forall j \neq i \mid W_i = u)$ is right continuous at x. The same conclusion clearly holds for $2 \leq i \leq p$. Therefore, we conclude that, for a.e. $x \in \mathbb{R}$, as $\epsilon \downarrow 0$,
$$\frac{1}{\epsilon}\mathrm{P}(A_1^{x,\epsilon}) \rightarrow \sum_{i=1}^p \mathrm{P}(W_j \leq x, \forall j \neq i \mid W_i = x)\phi(x - \mu_i)$$
$$= \phi(x) \sum_{i=1}^p e^{\mu_i x - \mu_i^2/2} \mathrm{P}(W_j \leq x, \forall j \neq i \mid W_i = x).$$

In the rest of the proof, we show that, for every $2 \leq i \leq p$ and $x \in \mathbb{R}$, $\mathrm{P}(A_i^{x,\epsilon}) = o(\epsilon)$ as $\epsilon \downarrow 0$, which leads to the desired conclusion. Fix any $2 \leq i \leq p$. The probability $\mathrm{P}(A_i^{x,\epsilon})$ is bounded by a sum of terms of

the form $P(x < W_j \le x + \epsilon, x < W_k \le x + \epsilon)$ with $j \neq k$. Recall that $\mathrm{Corr}(W_j, W_k) < 1$. Assume that $\mathrm{Corr}(W_j, W_k) = -1$. Then for every $x \in \mathbb{R}$, $P(x < W_j \le x + \epsilon, x < W_k \le x + \epsilon)$ is zero for sufficiently small ϵ. Otherwise, $(W_j, W_k)^T$ obeys a two-dimensional, non-degenerate Gaussian distribution and hence $P(x < W_j \le x + \epsilon, x < W_k \le x + \epsilon) = O(\epsilon^2) = o(\epsilon)$ as $\epsilon \downarrow 0$ for every $x \in \mathbb{R}$. This completes the proof. $\square$

Proof of Lemma 6. Since $E[W_0] \ge 0$, the map $x \mapsto \exp\{E[W_0]x - (E[W_0])^2/2\}$ is non-decreasing. Thus it suffices to show that the map

$$x \mapsto P(W_1 \le x, \dots, W_p \le x \mid W_0 = x) \tag{20}$$

is non-decreasing. As in the proof of Lemma 5, let $X_j = W_j - E[W_j]$ and let $V_j = X_j - E[X_j X_0]X_0$ be the residual from the orthogonal projection of X_j on X_0. Note that the vector $(V_j)_{1 \le j \le p}$ and X_0 are independent. Hence the probability in (20) equals

$$P(V_j \le x - \mu_j - E[X_j X_0](x - E[W_0]), 1 \le \forall j \le p \mid X_0 = x - E[W_0])$$
$$= P(V_j \le x - \mu_j - E[X_j X_0](x - E[W_0]), 1 \le \forall j \le p),$$

where the latter is non-decreasing in x on $\mathbb{R}$ since $E[X_j X_0] \le 1$. $\square$

APPENDIX A. PROOF OF LEMMA 1

Lemma 1 follows from the following maximal inequality and Hölder's inequality. Here we write $a \lesssim b$ if a is smaller than or equal to b up to a universal positive constant.

Lemma 8. *Let $Z_1, \dots, Z_n$ be independent random vectors in $\mathbb{R}^p$ with $p \ge 2$. Define $M := \max_{1 \le i \le n} \max_{1 \le j \le p} |Z_{ij}|$ and $\sigma^2 := \max_{1 \le j \le p} \sum_{i=1}^n E[Z_{ij}^2]$. Then*

$$E[\max_{1 \le j \le p} |\textstyle\sum_{i=1}^n (Z_{ij} - E[Z_{ij}])|] \lesssim (\sigma\sqrt{\log p} + \sqrt{E[M^2]}\log p).$$

We shall use the following lemma.

Lemma 9. *Let $V_1, \dots, V_n$ be independent random vectors in $\mathbb{R}^p$ with $p \ge 2$ such that $V_{ij} \ge 0$ for all $1 \le i \le n$ and $1 \le j \le p$. Then*

$$E[\max_{1 \le j \le p} \textstyle\sum_{i=1}^n V_{ij}] \lesssim \max_{1 \le j \le p} E[\textstyle\sum_{i=1}^n V_{ij}] + E[\max_{1 \le i \le n} \max_{1 \le j \le p} V_{ij}] \log p.$$

Proof of Lemma 9. We make use of the symmetrization technique. Let $\varepsilon_1, \dots, \varepsilon_n$ be independent Rademacher random variables (that is, $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = 1/2$) independent of $V_1^n := \{V_1, \dots, V_n\}$. Then by the triangle inequality and Lemma 2.3.1 in [28],

$$I := E[\max_{1 \le j \le p} \textstyle\sum_{i=1}^n V_{ij}] \le \max_{1 \le j \le p} E[\textstyle\sum_{i=1}^n V_{ij}] + E[\max_{1 \le j \le p} |\textstyle\sum_{i=1}^n (V_{ij} - E[V_{ij}])|]$$
$$\le \max_{1 \le j \le p} E[\textstyle\sum_{i=1}^n V_{ij}] + 2E[\max_{1 \le j \le p} |\textstyle\sum_{i=1}^n \varepsilon_i V_{ij}|].$$

By Lemmas 2.2.2 and 2.2.7 in [28], we have

$$\mathrm{E}[\max_{1\leq j\leq p}|\sum_{i=1}^n \varepsilon_i V_{ij}| \mid V_1^n] \lesssim \max_{1\leq j\leq p}(\sum_{i=1}^n V_{ij}^2)^{1/2}\sqrt{\log p}$$
$$\leq \sqrt{B\log p}\max_{1\leq j\leq p}(\sum_{i=1}^n V_{ij})^{1/2},$$

where $B := \max_{1\leq i\leq n}\max_{1\leq j\leq p} V_{ij}$. Hence by Fubini's theorem and the Cauchy-Schwarz inequality,

$$\mathrm{E}[\max_{1\leq j\leq p}|\sum_{i=1}^n \varepsilon_i V_{ij}|] \lesssim \sqrt{\mathrm{E}[B]\log p}(\mathrm{E}[\max_{1\leq j\leq p}\sum_{i=1}^n V_{ij}])^{1/2}$$
$$= \sqrt{\mathrm{E}[B]\log p}\sqrt{I}.$$

Therefore, we have

$$I \lesssim \max_{1\leq j\leq p}\mathrm{E}[\sum_{i=1}^n V_{ij}] + \sqrt{\mathrm{E}[B]\log p}\sqrt{I} =: a + b\sqrt{I}.$$

Solving this inequality, we conclude that $I \lesssim a + b^2$.　　　　□

Proof of Lemma 8. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables independent of $Z_1, \ldots, Z_n$. Then arguing as in the previous proof, we have

$$\mathrm{E}[\max_{1\leq j\leq p}|\sum_{i=1}^n (Z_{ij} - \mathrm{E}[Z_{ij}])|] \leq 2\mathrm{E}[\max_{1\leq j\leq p}|\sum_{i=1}^n \varepsilon_i Z_{ij}|]$$
$$\lesssim \mathrm{E}[\max_{1\leq j\leq p}(\sum_{i=1}^n Z_{ij}^2)^{1/2}]\sqrt{\log p}$$
$$\leq (\mathrm{E}[\max_{1\leq j\leq p}\sum_{i=1}^n Z_{ij}^2])^{1/2}\sqrt{\log p}. \quad \text{(Jensen)}$$

By Lemma 9 applied to $V_{ij} = Z_{ij}^2$, we have

$$\mathrm{E}[\max_{1\leq j\leq p}\sum_{i=1}^n Z_{ij}^2] \lesssim \sigma^2 + \mathrm{E}[M^2]\log p.$$

This implies the desired conclusion.　　　　□

Acknowledgments

References

[1] Adler, R. and Taylor, J. (2007). *Random Fields and Geometry.* Springer.
[2] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411-420.
[3] Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Infer.* **113** 385-402.

[4] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* **30** 205-216.

[5] Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods,Theory and Applications*. Springer.

[6] Candes, E. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313-2351.

[7] Chatterjee, S. (2005). An error bound in the Sudakov-Fernique inequality. arXiv:math/0510424.

[8] Chen, L., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. Springer.

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

[10] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013b). Anti-concentration and honest, adaptive confidence bands. arXiv:1303:7152.

[11] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

[12] Gordon, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265-289.

[13] Klivans, A.R., O'Donnell, R., and Servedio, R.A. (2008). Learning geometric concepts via Gaussian surface area. In: *Proc. 49th IEEE Symp. on Foundations of Comp. Science (FOCS)* pp. 541-550.

[14] Leadbetter, M., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer.

[15] Ledoux, M. (2001). *Concentration of Measure Phenomenon*. American Mathematical Society.

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

[17] Li, W. and Shao, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. In: *Handbook of Statistics Vol. 19*, pp. 533-597. North-Holland.

[18] Li, W. and Shao, Q.-M. (2002). A normal comparison inequality and applications. *Probab. Theory Relat. Fields* **122** 494-508.

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

[20] Nourdin, I. and Viens, F.G. (2009). Density formula and concentration inequalities with Malliavin calculus. *Electron. J. Probab.* **14** 2287-2309.

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

[22] Rudelson, M. and Vershynin, R. (2008). The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600-633.

[23] Rudelson, M. and Vershynin, R. (2009). Smallest singular value of a random rectangular matrix. *Commun. Pure Appl. Anal.* **62** 1707-1739.

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

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

[26] Sudakov, V.N. and Tsirel'son, B.S. (1978). Extremal properties of half-spaces for spherically invariant measures. *J. Soviet Math.* **9** 9-18.

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

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

[29] Vershynin, R. and Rudelson, M. (2007). Anti-concentration inequalities. Conference Proceeding for Phenomena in High Dimensions, Samos, Greece.

[30] Yan, L. (2009). Comparison inequalities for one sided normal probabilities. *J. Theoret. Probab.* **22** 827-836.

[31] Ylvisaker, D. (1965). The expected number of zeros of a stationary Gaussian process. *Ann. Math. Statist.* **36** 1043-1046.

[32] Ylvisaker, D. (1968). A note on the absence of tangencies in Gaussian sample paths. *Ann. Math. Statist.* **39** 261-262.

(V. Chernozhukov) Department of Economics & Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.

E-mail address: vchern@mit.edu

(D. Chetverikov) Department of Economics, UCLA, Bunche Hall, 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.

E-mail address: chetverikov@econ.ucla.edu

(K. Kato) Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail address: kkato@e.u-tokyo.ac.jp