

# A sharp lower-tail bound for Gaussian maxima with application to bootstrap methods in high dimensions\*

Miles E. Lopes and Junwen Yao

*University of California, Davis*  
 e-mail: [melopes@ucdavis.edu](mailto:melopes@ucdavis.edu); [jwyao@ucdavis.edu](mailto:jwyao@ucdavis.edu)

**Abstract:** Although there is an extensive literature on the maxima of Gaussian processes, there are relatively few non-asymptotic bounds on their lower-tail probabilities. The aim of this paper is to develop such a bound, while also allowing for many types of dependence. Let  $(\xi_1, \dots, \xi_N)$  be a centered Gaussian vector with standardized entries, whose correlation matrix  $R$  satisfies  $\max_{i \neq j} R_{ij} \leq \rho_0$  for some constant  $\rho_0 \in (0, 1)$ . Then, for any  $\epsilon_0 \in (0, \sqrt{1 - \rho_0})$ , we establish an upper bound on the probability  $\mathbb{P}(\max_{1 \leq j \leq N} \xi_j \leq \epsilon_0 \sqrt{2 \log(N)})$  in terms of  $(\rho_0, \epsilon_0, N)$ . The bound is also sharp, in the sense that it is attained up to a constant, independent of  $N$ . Next, we apply this result in the context of high-dimensional statistics, where we simplify and weaken conditions that have recently been used to establish near-parametric rates of bootstrap approximation. Lastly, an interesting aspect of this application is that it makes use of recent refinements of Bourgain and Tzafriri’s “restricted invertibility principle”.

**MSC2020 subject classifications:** Primary 60G15, 60E15; secondary 62G09, 62G32.

**Keywords and phrases:** Gaussian processes, tail bounds, high-dimensional statistics, bootstrap.

Received February 2021.

## 1. Introduction

The maxima of Gaussian processes play an essential role in many aspects of probability and statistics, and the literature describing them is highly developed [e.g. LLR83, Adl90, Lif95, LT13, Tal14]. Within this area, a variety of questions are related to showing that the maximum of a process is unlikely to deviate far above, or below, its mean. However, in comparison to the set of tools for handling the upper tail of a maximum, there are relatively few approaches for the lower tail. (Additional commentary related to this distinction may be found in [Tal14, p.viii] [LS01, Sec.4.2] [Lif95, Sec.18].)

In this paper, our goal is to derive lower-tail bounds for Gaussian maxima that are motivated by statistical applications involving bootstrap methods in high dimensions. We desire bounds that are general enough to handle many

---

\*Miles E. Lopes was supported in part by NSF grant DMS 1613218 and 1915786

types of correlation structures, yet also precise enough to yield explicit rates of convergence in distributional approximation results.

To describe the setting of our lower-tail bounds, let  $\xi = (\xi_1, \dots, \xi_N)$  be Gaussian vector with a correlation matrix  $R \in \mathbb{R}^{N \times N}$ , as well as  $\mathbb{E}(\xi_j) = 0$  and  $\text{var}(\xi_j) = 1$  for all  $1 \leq j \leq N$ . We will consider the situation where there is a fixed constant  $\rho_0 \in (0, 1)$  such that

$$\max_{i \neq j} R_{ij} \leq \rho_0. \quad (1.1)$$

In addition, we will consider some relaxations of this condition, where  $\rho_0 = 1$ , or where only subset of the off-diagonal entries of  $R$  are bounded above by a given constant. (See Corollary 2.1 and Theorem 2.3.) It is also worth noting that (1.1) does not require  $R$  to be invertible. Letting the maximum entry of  $\xi$  be denoted as

$$M_N(\xi) = \max_{1 \leq j \leq N} \xi_j,$$

we seek non-asymptotic upper bounds on the probability  $\mathbb{P}(M_N(\xi) \leq t)$ , where  $t$  is a suitable point in the lower tail.

### 1.1. Background

We now briefly review some leading results on lower-tail bounds for  $M_N(\xi)$ . Under the preceding conditions, the well-known concentration inequality for Lipschitz functions of Gaussian vectors implies that for any  $s > 0$ ,

$$\mathbb{P}\left(M_N(\xi) \leq \text{med}(M_N(\xi)) - s\right) \leq e^{-s^2/2}, \quad (1.2)$$

where  $\text{med}(\cdot)$  is any median [ST74, Bor75]. Although this bound is broadly applicable, it can fail to describe lower-tail probabilities smaller than  $\mathcal{O}(N^{-1})$ . To see this, consider using (1.2) to bound the probability  $\mathbb{P}(M_N(\xi) \leq \delta_0 \text{med}(M_N(\xi)))$  for some fixed  $\delta_0 \in (0, 1)$ . If the entries of  $\xi$  are independent, then the standard fact that  $\text{med}(M_N(\xi)) = \sqrt{2 \log(N)}(1 + o(1))$  as  $N \rightarrow \infty$  implies that (1.2) cannot give a bound better than  $\mathcal{O}(N^{-1})$  in this case. Furthermore, such a bound is far too large. In fact, when the entries of  $\xi$  are independent, this probability is  $\mathcal{O}(\exp(-aN^b))$ , for some fixed constants  $a, b > 0$  that may depend on  $\delta_0$  but not  $N$  [cf. Sch07, Section 2]. More recently, an important result of Paouris and Valettas [PV18] went beyond (1.2), showing that the inequality

$$\mathbb{P}\left(M_N(\xi) \leq \text{med}(M_N(\xi)) - s\right) \leq \frac{1}{2} \exp\left(-\frac{\pi}{1024} \frac{s^2}{\text{var}(M_N(\xi))}\right) \quad (1.3)$$

holds for any  $s > 0$ , which can improve upon (1.2) in situations where  $\text{var}(M_N(\xi))$  is small. For further variations and related results, we refer to the papers [PV19, Tan19, Val19].

Despite the progress achieved by the bound (1.3), it can still be quite challenging to obtain precise control on the variance  $\text{var}(M_N(\xi))$  in the exponent.

Indeed, it is often the case that bounds on  $\text{var}(M_N(\xi))$  have implicit dependence on the correlation matrix  $R$  that is difficult to quantify, and such bounds usually involve constants that are unspecified or conservative. (Note that since  $\text{var}(M_N(\xi))$  appears in the exponent, the constant in a variance bound will typically affect the rate of the tail bound.) We refer to the books [BLM13, Cha14] for further background on variance bounds related to  $M_N(\xi)$ .

**Contributions.** With regard to the considerations above, a few other works have developed lower-tail bounds for  $M_N(\xi)$  that provide insight into the role of the correlation structure [Har14, DEZ15, Tan15]. However, a limitation shared by all of these works is that they do not explicitly quantify the rate at which probabilities of the form  $\mathbb{P}(M_N(\xi) \leq \epsilon_0 \sqrt{2 \log(N)})$  decrease with respect to  $N$  for a fixed value  $\epsilon_0 \in (0, \sqrt{1 - \rho_0})$ . For instance, this limitation can arise from unspecified constants in the exponents of the bounds. See also the discussion after Theorem 2.1 below for more details.

In light of this, an important contribution of our lower-tail bounds is that they provide rates with explicit constants in their exponents. Also, our work goes further by showing that the rates are *sharp* (as in (2.3) and (2.4) of Theorem 2.1). Moreover, the proof techniques used here are entirely different from those used in the previously mentioned works. In particular, we extend and apply our lower-tail bounds by leveraging recent refinements of Bourgain and Tzafriri’s *restricted invertibility principle* (Section 2.1), which has been a topic of substantial interest in contemporary mathematics. Conventionally, this principle is understood as a functional-analytic result that guarantees the existence of special submatrices within large matrices. Hence, our use of this result to serve the quite different purpose of enhancing tail bounds may be viewed as another notable aspect of our work.

In addition to these contributions, Theorem 3.1 in Section 3 shows how our lower-tail bounds can be applied in the context of high-dimensional statistics, in order to simplify and weaken conditions that are sufficient for near-parametric rates of bootstrap approximation. Specifically, Theorem 3.1 advances the state of the art on such results for “max statistics” in settings where the data satisfy a condition known as “variance decay”. Also, a second application of our results has recently been developed by [Yi21] in connection with stochastic PDEs. Brief descriptions of both applications are given below.

## 1.2. Applications

**Bootstrap methods in high dimensions.** In recent years, inference problems related to max statistics have attracted significant attention in the high-dimensional statistics literature. The prototypical example of a max statistic is the coordinate-wise maximum of a sum of random vectors  $X_1, \dots, X_n \in \mathbb{R}^p$ , say

$$T = \max_{1 \leq j \leq p} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij}. \quad (1.4)$$

Statistics of this type arise often in the construction of hypothesis tests and confidence regions, and consequently, the performance of many inference procedures is determined by how accurately the distribution  $\mathcal{L}(T)$  can be approximated.

Bootstrap methods are a general approach to construct such approximations in a data-driven way, and they are designed to generate a random variable  $T^*$  whose conditional distribution given the data, denoted  $\mathcal{L}(T^*|X)$ , is close to  $\mathcal{L}(T)$ . In this context, accuracy is commonly measured with respect to the Kolmogorov metric  $d_K$ , and the challenge of developing non-asymptotic upper bounds on  $d_K(\mathcal{L}(T^*|X), \mathcal{L}(T))$  has stimulated an active line of work [e.g. CCKK19, CCK20, DZ20, LLM20, Lop20]. As an application of our lower-tail bounds, we will consider a recent bootstrap approximation result of this type from [LLM20], and we will improve upon it by showing that it holds under assumptions that are simpler and less restrictive.

**Macroscopic properties of solutions to stochastic PDEs.** In the study of stochastic partial differential equations, it is of interest to determine whether or not solutions exhibit high peaks over large regions at different scales. Solutions having this property are said to be “multifractal at macroscopic scales”. In order to demonstrate that a solution has this property in a precise sense, it is necessary to analyze regions where a solution rises above a certain height (exceedance sets), and quantify the “macroscopic Hausdorff dimension” of such regions.

During the last few years, a growing number of results have demonstrated that solutions to fundamental stochastic PDEs possess this multifractal property [e.g. KKK17, KKK18, Kim19, Yi21]. Quite recently, our first main result (Theorem 2.1) was used in [Yi21] to establish this property for certain versions of the stochastic heat and wave equations. Specifically, the lower-tail bound in our Theorem 2.1 was sharp enough to enable exact calculations of the macroscopic Hausdorff dimension of exceedance sets of certain Gaussian random fields. (See Theorems 1.1 and 1.3, as well as Section 3.2 in [Yi21].)

### 1.3. Notation

In order to simplify presentation, we always implicitly assume that  $N \geq 2$ , and we allow symbols for constants such as  $c, C, c_0, C_0, \dots$  to change values with each appearance. When dealing with iterated logarithms, we use the abbreviation  $\ell_2(N) := \max\{1, \log \log(N)\}$ . For a real matrix  $A$ , define  $\|A\|_F = \sqrt{\text{tr}(A^\top A)}$ ,  $\|A\|_1 = \sum_{i,j} |A_{ij}|$ , and also define  $\|A\|_{\text{op}}$  to be the largest singular value of  $A$ . For a real vector  $x$ , we write  $\|x\|_2$  for the Euclidean norm. If  $A \in \mathbb{R}^{N \times N}$  is symmetric, its sorted eigenvalues are  $\lambda_{\max}(A) = \lambda_1(A) \geq \dots \geq \lambda_N(A) = \lambda_{\min}(A)$ . The identity matrix of size  $N \times N$  is  $I_N$ , and the standard basis vectors in  $\mathbb{R}^N$  are  $e_1, \dots, e_N$ . For the distribution of a random variable  $V$ , we write  $\mathcal{L}(V)$ , and we define its  $\psi_1$ -Orlicz norm as  $\|V\|_{\psi_1} = \inf\{t > 0 \mid \mathbb{E}[\exp(|V|/t)] \leq 2\}$ . If  $\zeta$  is a Gaussian random vector with mean 0 and covariance matrix  $\Sigma$ , we write  $\zeta \sim \mathcal{N}(0, \Sigma)$ . For the univariate standard Gaussian distribution, the symbols  $\Phi$

and  $\phi$  denote the distribution function and density. Lastly, if  $a$  and  $b$  are real numbers, we use the notation  $a \vee b = \max\{a, b\}$  and  $a \wedge b = \min\{a, b\}$ .

## 2. Lower-tail bounds

To clarify the statements of our main results, we first state a basic proposition describing the sizes of  $\mathbb{E}(M_N(\xi))$  and  $\text{med}(M_N(\xi))$  under condition (1.1). This proposition shows that the value  $\sqrt{2(1-\rho_0)\log(N)}$  is a natural reference level for a lower-tail bound. Although this fact might be considered well-known among specialists, it is not easily referenced in the form given below, and so we provide a short proof at the end of Section 4.

**Proposition 2.1.** *Let  $\mu_N$  stand for either  $\mathbb{E}(M_N(\xi))$  or  $\text{med}(M_N(\xi))$ . If the condition (1.1) holds for some  $\rho_0 \in (0, 1)$ , then there is a universal constant  $c_0 > 0$  such that*

$$\mu_N \geq \sqrt{2(1-\rho_0)\log(N)} - c_0\sqrt{\ell_2(N)}. \quad (2.1)$$

Furthermore, if  $R_{ij} = \rho_0$  for all  $i \neq j$ , then

$$\mu_N \leq \sqrt{2(1-\rho_0)\log(N)}. \quad (2.2)$$

The following is our first main result, which will be extended and refined later in Corollary 2.1 and Theorem 2.3. The proof is deferred to Section 4.

**Theorem 2.1.** *Fix two constants  $\delta_0, \rho_0 \in (0, 1)$  with respect to  $N$ , and suppose the condition (1.1) holds. Then, there is a constant  $C > 0$  depending only on  $(\delta_0, \rho_0)$ , such that*

$$\mathbb{P}\left(M_N(\xi) \leq \delta_0\sqrt{2(1-\rho_0)\log(N)}\right) \leq C N^{\frac{-(1-\rho_0)(1-\delta_0)^2}{\rho_0}} (\log(N))^{\frac{1-\rho_0(2-\delta_0)-\delta_0}{2\rho_0}}. \quad (2.3)$$

Furthermore, the bound (2.3) is sharp in the sense that if  $R_{ij} = \rho_0$  for all  $i \neq j$ , then there is a constant  $c > 0$  depending only on  $(\delta_0, \rho_0)$ , such that

$$\mathbb{P}\left(M_N(\xi) \leq \delta_0\sqrt{2(1-\rho_0)\log(N)}\right) \geq c N^{\frac{-(1-\rho_0)(1-\delta_0)^2}{\rho_0}} (\log(N))^{\frac{1-\rho_0(2-\delta_0)-\delta_0}{2\rho_0}}. \quad (2.4)$$

**Remarks.** To comment on some basic features of the theorem, first note that the dominant exponent  $-(1-\rho_0)(1-\delta_0)^2/\rho_0$  takes larger negative values as  $\rho_0$  becomes smaller. Hence, the bound respects the fact that the lower-tail probability decays faster than any power of  $N^{-1}$  when the entries of  $\xi$  are independent. Also, the theorem conforms with the reference level motivated by Proposition 2.1, since (2.3) implies that  $\text{med}(M_N(\xi))$  cannot be much less than  $\sqrt{2(1-\rho_0)\log(N)}$ .

Regarding other recent lower-tail bounds for  $M_N(\xi)$ , their relation to (2.3) and (2.4) can be summarized as follows (in the setting of Theorem 2.1 with  $\rho_0$

and  $\delta_0$  regarded as fixed with respect to  $N$ ). First, the paper [DEZ15, Theorem 1.6] gives a two-sided tail bound for  $M_N(\xi)$ , implying that if  $a$  and  $b$  are positive constants satisfying  $\mathbb{E}(M_N(\xi)) \geq a\sqrt{\log(N)}$  and  $b \leq a/100$ , then the probability  $\mathbb{P}(M_N(\xi) \leq \mathbb{E}(M_N(\xi)) - b\sqrt{\log(N)})$  is at most of order  $N^{-b^2/(2-c(a))}$ , for an unspecified constant  $c(a) > 0$ . Second, the paper [Tan15, Proposition 7] gives a two-sided tail bound implying that for any constant  $b > 0$ , the probability  $\mathbb{P}(M_N(\xi) \leq \mathbb{E}(M_N(\xi)) - b\sqrt{\log(N)})$  is at most of order  $\exp(-\frac{cb}{\sqrt{\rho_0}}\sqrt{\log(N)})$ , where  $c > 0$  is an unspecified constant. (Note that even if  $c$  were specified, this bound would be of larger order than  $N^{-\epsilon}$  for any  $\epsilon \in (0, 1)$ .) Third, the paper [Har14, Theorem 3.4] develops a lower-tail bound of the form  $\mathbb{P}(M_N(\xi) \leq t_N(\alpha)) \leq 2\alpha$  for any  $\alpha \in (0, 1/2)$ , where  $t_N(\alpha)$  is a threshold with a complex dependence on  $N$ ,  $\alpha$ , as well as other parameters related to the correlation structure of  $\xi$ . In particular, Section 4 of the paper explains that the threshold can be expressed in terms of the minimum eigenvalue of  $R$ . For instance, when  $R$  satisfies  $R_{ij} = \rho_0$  for all  $i \neq j$ , the minimum eigenvalue is equal to  $1 - \rho_0$ , and in this case, Theorem 3.4 and the comments preceding equation (4.1) in [Har14] yield the following formula:  $t_N(\alpha) = \sqrt{1 - \rho_0}[\kappa_N(\alpha) - \sqrt{\log(\kappa_N(\alpha))} + \sqrt{\rho_0}\Phi^{-1}(\alpha)]$ , where we put  $\kappa_N(\alpha) = \log(N^2/(2\pi)) - 2\log(-\log(\alpha))$ . However, it seems that this formula for  $t_N(\alpha)$  is not quite correct, since it is missing a square root. (Note that the quantity  $\kappa_N(\alpha)$  scales like  $\log(N)$ , rather than  $\sqrt{\log(N)}$ , as a function of  $N$ .) Apart from this issue, the intricate form of the threshold also makes our result ostensibly easier to use. Lastly, in comparison to our work, none of the three mentioned papers resolve the question of whether or not the lower-tail bounds yield sharp rates.

**Extra correlation structure.** We now present a direct corollary of Theorem 2.1 that allows for extra structure in the matrix  $R$  to be used. If the matrix  $R$  satisfies  $\max_{i \neq j} R_{ij} = \rho_0$ , but “most” off-diagonal entries are substantially less than  $\rho_0$ , then we can gain considerable improvement. Roughly speaking, if there is a number  $\rho_1 \in (0, \rho_0)$  such that a large number of off-diagonal entries satisfy  $R_{ij} \leq \rho_1$ , then Theorem 2.1 can be improved by effectively replacing  $\rho_0$  with the better value  $\rho_1$ .

**Corollary 2.1.** *Fix two constants  $\delta_1, \rho_1 \in (0, 1)$  with respect to  $N$ . Suppose there is an index set  $J \subset \{1, \dots, N\}$  with cardinality  $|J| \geq 2$ , such for any distinct  $i, j \in J$ , the bound  $R_{ij} \leq \rho_1$  holds. Then, there is a constant  $C > 0$  depending only on  $(\delta_1, \rho_1)$  such that*

$$\mathbb{P}\left(M_N(\xi) \leq \delta_1 \sqrt{2(1 - \rho_1) \log(|J|)}\right) \leq C |J|^{\frac{-(1-\rho_1)(1-\delta_1)^2}{\rho_1}} (\log(|J|))^{\frac{1-\rho_1(2-\delta_1)-\delta_1}{2\rho_1}}.$$

Another point worth noting about this corollary is that it remains applicable even in situations where some variables are perfectly correlated and  $\rho_0 = 1$ .

### 2.1. Further improvements using the restricted invertibility principle

In order to leverage the full strength of Corollary 2.1, we need an index set  $J \subset \{1, \dots, N\}$  with large cardinality, such that the off-diagonal entries  $R_{ij}$  with  $i, j \in J$  are uniformly small. Quite remarkably, it turns out that such an index set is guaranteed to exist under rather general conditions, as a consequence of a seminal result known as the *restricted invertibility principle* [BT87]. Over the past decade, this result has received considerable attention, due to its close links with the solution of the longstanding Kadison-Singer problem by Marcus, Spielman and Srivastava [MSS15]. Below, we will apply a recent refinement of the restricted invertibility principle, established in a companion paper by the same group [MSS21].

To introduce some notation, define the *stable rank*  $\mathbf{r}(A)$  of a non-zero positive semidefinite matrix  $A$  as

$$\mathbf{r}(A) = \frac{[\text{tr}(A)]^2}{\|A\|_F^2}.$$

This quantity always satisfies

$$1 \leq \mathbf{r}(A) \leq \text{rank}(A),$$

and has the useful property that it approximately counts the number of dominant eigenvalues of  $A$ . In terms of the notion of stable rank, the restricted invertibility principle roughly says that for any matrix  $L \in \mathbb{R}^{N \times N}$ , there exists an index set  $J \subset \{1, \dots, N\}$  with cardinality  $|J| \approx \mathbf{r}(L^\top L)$ , such that the column submatrix of  $L$  indexed by  $J$  has (non-trivial) singular values that are well separated from zero.

**Theorem 2.2** (Restricted invertibility principle [MSS21]). *Let  $L \in \mathbb{R}^{N \times N}$  be a non-zero matrix. Then, for any positive integer  $k \leq \mathbf{r}(L^\top L)$ , there exists an index set  $J \subset \{1, \dots, N\}$  with cardinality  $|J| = k$ , such that the following inequality holds for any real numbers  $(a_j)_{j \in J}$ ,*

$$\left\| \sum_{j \in J} a_j L e_j \right\|_2^2 \geq \frac{\|L\|_F^2}{N} \cdot \left(1 - \sqrt{\frac{k}{\mathbf{r}(L^\top L)}}\right)^2 \cdot \sum_{j \in J} a_j^2. \quad (2.5)$$

We will apply the restricted invertibility principle above with  $R^{1/2}$  playing the role of  $L$ . Note that because  $R$  is a correlation matrix, we have  $\|R^{1/2}\|_F^2/N = 1$  and  $\mathbf{r}(R) = N^2/\|R\|_F^2$ . To simplify the inequality (2.5), let  $R(J) \in \mathbb{R}^{|J| \times |J|}$  denote the submatrix of  $R$  with entries indexed by  $J \times J$ . Also, suppose there is an integer  $k \geq 2$  and a scalar  $\epsilon \in (0, 1)$ , such that  $k \leq \frac{\epsilon^2}{4} \mathbf{r}(R)$ . In this case, the restricted invertibility principle ensures there is a choice of  $J$  with cardinality equal to  $k$  such that

$$\lambda_{\min}(R(J)) \geq (1 - \epsilon/2)^2.$$

In turn, this implies that the off-diagonal entries of  $R(J)$  are uniformly small, i.e.

$$\begin{aligned} \max_{i \neq j} R_{ij}(J) &= 1 - \min_{i \neq j} \frac{1}{2} (e_i - e_j)^\top R(J) (e_i - e_j) \\ &\leq 1 - \lambda_{\min}(R(J)) \\ &\leq \epsilon. \end{aligned}$$

The next theorem combines this information about  $J$  with Corollary 2.1. Later on, this connection will be used in our application dealing with rates of bootstrap approximation in high dimensions.

**Theorem 2.3.** *Fix two constants  $\delta, \epsilon \in (0, 1)$  with respect to  $N$ , and consider a Gaussian random vector  $\xi \sim \mathcal{N}(0, R)$  for some correlation matrix  $R \in \mathbb{R}^{N \times N}$ . Then, there is a constant  $C > 0$  depending only on  $(\delta, \epsilon)$  such that the following inequality holds for any integer  $k \geq 2$  satisfying  $k \leq \frac{\epsilon^2}{4} \mathbf{r}(R)$ ,*

$$\mathbb{P}\left(M_N(\xi) \leq \delta \sqrt{2(1-\epsilon) \log(k)}\right) \leq C k^{\frac{-(1-\epsilon)(1-\delta)^2}{\epsilon}} (\log(k))^{\frac{1-\epsilon(2-\delta)-\delta}{2\epsilon}}. \quad (2.6)$$

**Remarks.** To discuss the choice of  $k$ , consider a basic situation where the root mean square of eigenvalues, say  $\lambda_{\text{rms}}(R) = \frac{1}{\sqrt{N}} \sqrt{\lambda_1(R)^2 + \dots + \lambda_N(R)^2}$ , is upper bounded by a constant  $C$ . For instance, this condition is quite natural in the context of principal components analysis, where the matrix  $R$  may only have a handful of dominant eigenvalues. Note too that this condition is much weaker than requiring  $\|R\|_{\text{op}} \leq C$ , since it is possible for the largest eigenvalue of  $R$  to diverge while  $\lambda_{\text{rms}}(R)$  remains bounded as  $N \rightarrow \infty$ . Due to the fact that  $R$  is a correlation matrix, the condition  $\lambda_{\text{rms}}(R) \leq C$  implies

$$\mathbf{r}(R) \geq \frac{N}{C^2},$$

and consequently, the integer  $k$  in Theorem 2.3 may be chosen to be of order  $N$ . This leads to an upper bound in (2.6) of order  $N^{-(1-\epsilon)(1-\delta)^2/\epsilon}$ , up to a logarithmic factor. To see why this illustrates the benefit of the restricted invertibility principle, note that it enables us to bypass an assumption of the form  $\max_{i \neq j} R_{ij} \leq \epsilon$ , which would have been necessary if Theorem 2.1 had been used directly. Indeed, a condition like  $\lambda_{\text{rms}}(R) \leq C$  is often more appealing from a statistical standpoint, because it allows for many off-diagonal entries of  $R$  to be close to 1.

### 3. Application to bootstrap methods in high dimensions

**Preliminaries.** Let  $X_1, \dots, X_n \in \mathbb{R}^p$  be centered i.i.d. observations with a standardized sum denoted as  $S_n = n^{-1/2} \sum_{i=1}^n X_i$ . In addition, define the



coordinate-wise variances  $\sigma_j^2 = \text{var}(X_{1j})$  for each  $j = 1, \dots, p$ , which are assumed to be positive, and define the partially standardized max statistic

$$M = \max_{1 \leq j \leq p} S_{n,j} / \sigma_j^\tau,$$

where  $\tau \in [0, 1)$  is a tuning parameter. This type of max statistic was studied previously in [LLM20], and reduces to the basic example (1.4) in the case when  $\tau$  is chosen to be 0.

In order to analyze bootstrap approximations of  $\mathcal{L}(M)$ , we will focus on the “Gaussian multiplier bootstrap method” popularized in [CCK13]. To describe the method, define the sample mean  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , the sample covariance matrix  $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$ , and the coordinate-wise sample variances  $\hat{\sigma}_j^2 = \hat{\Sigma}_{jj}$  for  $j = 1, \dots, p$ . The bootstrapped statistic  $M^*$  is constructed by generating a Gaussian random vector  $S_n^* \sim \mathcal{N}(0, \hat{\Sigma})$ , and defining<sup>1</sup>

$$M^* = \max_{1 \leq j \leq p} S_{n,j}^* / \hat{\sigma}_j^\tau.$$

Here, the key point is that it is possible in practice to directly simulate the conditional distribution of  $M^*$  given the data, denoted  $\mathcal{L}(M^*|X)$ . Accordingly, the multiplier bootstrap method uses  $\mathcal{L}(M^*|X)$  as an approximation to  $\mathcal{L}(M)$ .

As one last preliminary item, we will adopt the standard convention in high-dimensional statistics of considering a sequence of models indexed by  $n$ , in which all parameters may depend on  $n$ , except when stated otherwise. In particular, the dimension  $p = p(n)$  is allowed to have arbitrary dependence on  $n$ . Likewise, if a parameter does not depend on  $n$ , then it does not depend on  $p$  either. To simplify notation, we will write  $a_n \lesssim b_n$  for two sequences of real numbers  $a_n$  and  $b_n$  if there is a constant  $c > 0$  not depending on  $n$  such that  $a_n \leq cb_n$  holds for all large  $n$ . In the case when both  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$  hold, we write  $a_n \asymp b_n$ .

### 3.1. A brief review of bootstrap approximation under variance decay

Recently, the paper [LLM20] analyzed how well  $\mathcal{L}(M^*|X)$  approximates  $\mathcal{L}(M)$  in the setting of “variance decay”, where the sorted coordinate-wise variances  $\sigma_{(1)}^2 \geq \dots \geq \sigma_{(p)}^2$  have a polynomial decay profile. This means that there is a constant  $\gamma > 0$  not depending on  $n$  such that the condition  $\sigma_{(j)} \asymp j^{-\gamma}$  holds for all  $j = 1, \dots, p$ . (The constant  $\gamma$  is allowed to be arbitrarily large or small.) For instance, the structure of variance decay arises naturally in a variety of high-dimensional statistical applications related to *principal components analysis*, *count data*, and the *Fourier coefficients of functional data*. Under the assumption of variance decay, as well as some additional assumptions on the correlation and tail-behavior of the covariates, Theorem 3.2 in the paper [LLM20] established the following bootstrap approximation result. Namely, for any fixed  $\delta \in (0, 1/2)$ , there is a constant  $C > 0$  not depending on  $n$  such that the bound

<sup>1</sup>The expression  $S_{n,j}^* / \hat{\sigma}_j^\tau$  is regarded as 0 in the exceptional case  $\hat{\sigma}_j = 0$ .

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(M^* \leq t | X) - \mathbb{P}(M \leq t) \right| \leq C n^{-\frac{1}{2} + \delta} \quad (3.1)$$

holds with probability at least  $1 - C/n$ .

The result (3.1) has some significant distinguishing features in relation to the other recent works [Lop20, CCK20] that have obtained near  $n^{-1/2}$  rates of bootstrap approximation for max statistics in high-dimensional settings. First, the use of variance decay makes it possible for the bound (3.1) to hold independently of  $p$ , whereas the other works do not use this structure and obtain rates with polylogarithmic dependence on  $p$ . Second, the other works require that the correlation matrix  $\text{cor}(X_1) \in \mathbb{R}^{p \times p}$  be either positive definite or well-approximated by a positive definite matrix, whereas the bound (3.1) does not. Nevertheless, the bound (3.1) does rely on some other assumptions about the matrix  $\text{cor}(X_1)$ . Below, we outline these assumptions, and later in Section 3.2, we will show how these assumptions can be simplified and weakened by applying our work from Section 2.

To introduce some notation, consider an arbitrary index  $d \in \{1, \dots, p\}$ , and let  $J(d) \subset \{1, \dots, p\}$  denote a set of indices corresponding to the  $d$  largest coordinate-wise variances, i.e.  $\{\sigma_{(1)}^2, \dots, \sigma_{(d)}^2\} = \{\sigma_j^2 \mid j \in J(d)\}$ . Next, let  $R(d)$  denote the  $d \times d$  correlation matrix of the variables  $(X_{1j})_{j \in J(d)}$ , and let the matrix  $R^+(d) \in \mathbb{R}^{d \times d}$  be the non-negative part,  $R_{ij}^+(d) = R_{ij}(d) \vee 0$ . Lastly, define the integer  $m_n = \lceil (\log(n))^3 \wedge p \rceil$ , which always satisfies  $1 \leq m_n \leq p$ .

In terms of this notation, the paper [LLM20] makes the following three correlation assumptions in order to establish (3.1):

- (a) There is a constant  $\epsilon_0 \in (0, 1)$  not depending on  $n$  such that

$$\max_{i \neq j} R_{ij}(m_n) \leq 1 - \epsilon_0.$$

- (b) The matrix  $R^+(m_n)$  is positive semidefinite.

- (c) The condition  $\|R(m_n)\|_1 \lesssim m_n$  holds.

Concerning assumption (b), this is non-trivial because the operation of taking the entrywise non-negative part of a matrix does not generally preserve positive semidefiniteness [GR15, Theorem 4.11].

### 3.2. Bootstrap approximation result with relaxed correlation assumptions

In this subsection, we will present a result of the form (3.1) in which the previous correlation assumptions (a) and (b) are removed, and in which (c) will be replaced by a version involving a weaker norm. Apart from the correlation structure, the following conditions on the data-generating model and the variance decay profile are slightly simplified versions of the ones used in [LLM20].

**Assumption 3.1** (Data-generating model and variance decay).

- (i) There is a positive semidefinite matrix  $\Sigma \in \mathbb{R}^{p \times p}$ , such that the observations  $X_1, \dots, X_n \in \mathbb{R}^p$  are generated as

$$X_i = \Sigma^{1/2} Z_i$$

for each  $i = 1, \dots, n$ , where  $Z_1, \dots, Z_n$  are i.i.d. random vectors, with  $\mathbb{E}(Z_1) = 0$ ,  $\mathbb{E}(Z_1 Z_1^\top) = I_p$ , as well as  $\sup_{\|u\|_2=1} \|Z_1^\top u\|_{\psi_1} \lesssim 1$ .

- (ii) The parameters  $\sigma_{(1)} \geq \dots \geq \sigma_{(p)}$  are positive, and there is a constant  $\gamma > 0$  not depending on  $n$  such that the condition

$$\sigma_{(j)} \asymp j^{-\gamma}$$

holds for all  $j = 1, \dots, p$ .

The following is the main result of the current section, and the proof will be given in Section 5.

**Theorem 3.1** (Bootstrap approximation). *Fix any constants  $\delta \in (0, 1/2)$  and  $\tau \in [0, 1)$  with respect to  $n$ , and define  $\kappa = 4 \vee (3\gamma(1 - \tau))$  as well as  $l_n = \lceil n^{\delta/\kappa} \wedge p \rceil$ . In addition, suppose that Assumption 3.1 holds, and that the condition*

$$\|R(l_n)\|_F^2 \lesssim l_n^{2-\delta}. \quad (3.2)$$

*holds. Then, there is a constant  $C > 0$  not depending on  $n$  such that the event*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(M^* \leq t | X) - \mathbb{P}(M \leq t) \right| \leq C n^{-\frac{1}{2} + \delta} \quad (3.3)$$

*holds with probability at least  $1 - C/n$ .*

**Remarks.** To interpret the correlation assumption (3.2), it should be emphasized that the upper bound  $\|R(l_n)\|_F^2 \leq l_n^2$  holds under all circumstances, because any positive semidefinite matrix  $A$  satisfies  $\|A\|_F^2 \leq [\text{tr}(A)]^2$ . So, given that  $\delta$  may be taken to be arbitrarily small, it is not possible to substantially weaken (3.2) in general. Furthermore, it is worth noting that the condition (3.2) only applies to the small set of variables indexed by  $J(l_n)$ , while all other variables indexed by  $\{1, \dots, p\} \setminus J(l_n)$  have a correlation structure that is unrestricted.

A large class of examples of correlation matrices satisfying (3.2) can be constructed in the following way. Let  $f : [0, \infty) \rightarrow [0, 1]$  be any continuous convex function that satisfies  $f(0) = 1$ , as well as  $f(t) \leq ct^{-\delta}$  for some fixed constants  $c > 0$ ,  $\delta \in (0, 1/2)$ , and all  $t \geq 0$ . Then, by Pólya's criterion [Pól49, Theorem 1], any matrix with entries defined by  $R_{ij} = f(|i - j|)$  will be a correlation matrix that satisfies (3.2). Moreover, any other correlation matrix obtained by permuting the rows and columns will continue to satisfy (3.2). For additional discussion of such matrices, and some of their connections to high-dimensional statistics, we refer to [BL08].

With regard to our lower-tail bounds from Section 2, the connection with Theorem 3.1 can be explained as follows. Overall, the proof of this result is based

on a dimension-reduction strategy that involves showing  $M$  is well approximated by a statistic of the form  $M' = \max_{j \in J'} S_{n,j}/\sigma_j^\tau$ , where  $J' \subset \{1, \dots, p\}$  is an index set with cardinality much less than  $p$ . In order to show that  $M = M'$  holds with high probability, we will localize the maximizing index for  $M$  to the set  $J'$ . That is, if  $\hat{j}$  is a random index such that  $M = S_{n,\hat{j}}/\sigma_{\hat{j}}^\tau$ , then we will show that  $\hat{j}$  falls into  $J'$  with high probability. It is in this localization step that the lower-tail bound from Theorem 2.3 will be used: Two values  $x < y$  will be determined, such that  $M'$  is likely to be above  $y$ , while the maximum of  $S_{n,j}/\sigma_j^\tau$  over the indices  $j \notin J'$  is likely to be below  $x$ . Hence, the first of these items requires a lower-tail bound, because it involves showing that the probability  $\mathbb{P}(M' \leq y)$  is small.

#### 4. Proof of Theorem 2.1

Throughout the proofs, we always assume that  $N$  is sufficiently large for any given expression to make sense — since this can be accommodated by an adjustment of the constants  $C$  and  $c$  in the statement of Theorem 2.1. The symbols  $c$  and  $C$  will also be frequently be re-used with the understanding that they never depend on  $N$  (and similarly with respect to the sample size  $n$  in Section 5). In addition, it will simplify presentation to introduce the quantities

$$\alpha_0 = \frac{(1-\rho_0)(1-\delta_0)^2}{\rho_0} \quad \text{and} \quad \beta_0 = \frac{(1-\rho_0)(1-\delta_0)}{\rho_0},$$

so that the bounds in Theorem 2.1 are proportional to  $N^{-\alpha_0}(\log(N))^{\frac{\beta_0-1}{2}}$ .

**Proof of the upper bound (2.3).** Let  $t_N = \delta_0 \sqrt{2(1-\rho_0)\log(N)}$ , and let  $\xi' \sim \mathcal{N}(0, R')$  be a Gaussian vector in  $\mathbb{R}^N$ , where  $R'$  is a correlation matrix satisfying  $R'_{ij} = \rho_0$  for all  $i \neq j$ . Due to the assumption that  $\max_{i \neq j} R_{ij} \leq \rho_0$ , it follows from Slepian's Lemma [Sle62] that

$$\mathbb{P}(M_N(\xi) \leq t_N) \leq \mathbb{P}(M_N(\xi') \leq t_N).$$

To control the larger probability, let  $\zeta_0, \zeta_1, \dots, \zeta_N$  denote independent  $\mathcal{N}(0, 1)$  random variables, and note that the coordinates of  $\xi'$  may be jointly represented in distribution as

$$\xi'_j = \sqrt{\rho_0}\zeta_0 + \sqrt{1-\rho_0}\zeta_j. \quad (4.1)$$

This yields the following representation of the maximum

$$M_N(\xi') = \sqrt{\rho_0}\zeta_0 + \sqrt{1-\rho_0} \max_{1 \leq j \leq N} \zeta_j, \quad (4.2)$$

which allows us to view  $M_N(\xi')$  as the convolution of the two independent random variables on the right side. Consequently, a direct calculation may be used to obtain the exact formula

$$\mathbb{P}(M_N(\xi') \leq t_N) = \int_{-\infty}^{\infty} \psi_N(s) ds, \quad (4.3)$$

where the integrand is defined by

$$\psi_N(s) = \phi(s) \Phi^N\left(\frac{t_N - \sqrt{\rho_0} s}{\sqrt{1 - \rho_0}}\right),$$

with  $\phi$  and  $\Phi$  being the standard normal density and distribution function. The remainder of the proof consists in bounding integral  $\int_{-\infty}^{\infty} \psi_N(s) ds$  over the intervals  $(-\infty, -c_N]$ ,  $[-c_N, 0]$ , and  $[0, \infty)$ , where we define

$$c_N = \sqrt{\frac{2(1-\rho_0)\log(N)}{\rho_0}} \left\{ 1 - \delta_0 - \frac{\frac{1}{4}\ell_2(N)}{\log(N)} \right\}. \quad (4.4)$$

*Remarks.* Our choice of the cut-off point  $c_N$  is a crucial element of the proof. To give a sense of how delicate this is, a close inspection of the proof shows that the  $1/4$  coefficient on the trailing term  $\ell_2(N)/\log(N)$  is needed to obtain both the upper and lower bounds in Theorem 2.1. Some intuition for the definition of  $c_N$  can be given as follows. First, define  $b_N = \sqrt{2\log(N)} - \ell_2(N)/\sqrt{8\log(N)}$  and the random variable  $G_N = \sqrt{2\log(N)}(\max_{1 \leq j \leq N} \zeta_j - b_N)$ , which has the property that  $G_N + \log(\sqrt{4\pi})$  converges weakly to a standard Gumbel distribution as  $N \rightarrow \infty$  [LLR83, Theorem 1.5.3]. Based on this notation and (4.2), it can be checked that

$$\mathbb{P}(M_N(\xi') \leq t_N) = \mathbb{P}\left(\zeta_0 + \sqrt{\frac{1-\rho_0}{2\rho_0\log(N)}} G_N \leq -c_N\right). \quad (4.5)$$

Next, the probability on the right can be approximated heuristically by dropping the random variable  $\sqrt{(1-\rho_0)/(2\rho_0)} G_N / \sqrt{\log(N)}$ , because it is of order  $1/\sqrt{\log(N)}$  (in probability), and hence asymptotically negligible compared to  $\zeta_0$ . After this heuristic is used, some further calculation then leads to the surmise that  $\mathbb{P}(M_N(\xi') \leq t_N)$  should be of order  $N^{-\alpha_0} \log(N)^{\frac{\beta_0-1}{2}}$ . However, this line of reasoning will not be used in the formal proof, because the bounds in Theorem 2.1 are non-asymptotic. Also, note that this asymptotic heuristic suppresses the fact that  $1/\sqrt{\log(N)}$  approaches 0 very slowly, which matters much more from a non-asymptotic standpoint.

Returning to the main thread, the problem of upper bounding the integral of  $\psi_N(s)$  over the interval  $[-c_N, 0]$  is the most involved part of the proof, and is postponed to Lemma 4.1 later on. (The reason for this difficulty is that the function  $\psi_N(s)$  is especially sensitive to small changes in  $s$  over  $[-c_N, 0]$ . In particular, this stage of the analysis involves developing separate bounds over two sub-intervals of  $[-c_N, 0]$  in order to account for changes in the local behavior of the function.) Once the proof of that lemma is complete, it will be straightforward to control the integral over  $[0, \infty)$ , which is done subsequently in Lemma 4.2. For the moment, we only handle the interval  $(-\infty, -c_N]$ , since

it requires no further preparation. Indeed, we have

$$\begin{aligned}
\int_{-\infty}^{-c_N} \psi_N(s) ds &\leq \int_{-\infty}^{-c_N} \phi(s) ds \\
&\leq \frac{e^{-c_N^2/2}}{\sqrt{2\pi c_N}} \quad (\text{Mill's inequality}) \\
&\leq \frac{C}{\sqrt{\log(N)}} N^{-\alpha_0} \exp\left\{\frac{\beta_0}{2} \ell_2(N)\right\} \\
&\leq C N^{-\alpha_0} (\log(N))^{\frac{\beta_0-1}{2}},
\end{aligned} \tag{4.6}$$

where  $C$  only depends on  $\delta_0$  and  $\rho_0$ . This proves the upper bound (2.3).

**Proof of the lower bound (2.4).** It suffices to derive a lower bound on the integral  $\int_{-\infty}^{-c_N} \psi_N(s) ds$ . Using the fact that  $\Phi^N\left(\frac{t_N - \sqrt{\rho_0} s}{\sqrt{1-\rho_0}}\right)$  is decreasing in  $s$ , we have

$$\int_{-\infty}^{-c_N} \psi_N(s) ds \geq \Phi^N\left(\frac{t_N + \sqrt{\rho_0} c_N}{\sqrt{1-\rho_0}}\right) \Phi(-c_N). \tag{4.7}$$

If we define the number

$$\begin{aligned}
r_N &= \frac{t_N + \sqrt{\rho_0} c_N}{\sqrt{1-\rho_0}} \\
&= \sqrt{2 \log(N)} \left\{1 - \frac{\frac{1}{4} \ell_2(N)}{\log(N)}\right\},
\end{aligned} \tag{4.8}$$

then Mill's inequality gives

$$\begin{aligned}
\Phi(r_N) &\geq 1 - \frac{1}{\sqrt{2\pi} r_N} e^{-r_N^2/2} \\
&\geq 1 - \frac{C}{N}.
\end{aligned} \tag{4.9}$$

Hence, the limit  $(1 - C/N)^N \rightarrow \exp(-C)$  as  $N \rightarrow \infty$  shows that  $\Phi^N(r_N)$  is lower bounded by a positive constant for all large  $N$ . Finally, regarding the factor  $\Phi(-c_N)$  in (4.7), the lower-bound version Mill's inequality [Dur05, Theorem 1.4] gives

$$\begin{aligned}
\Phi(-c_N) &\geq \left\{\frac{1}{c_N} - \frac{1}{c_N^3}\right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} c_N^2} \\
&\geq c N^{-\alpha_0} (\log(N))^{\frac{1}{2}(\beta_0-1)},
\end{aligned} \tag{4.10}$$

where the second step follows from a calculation very similar to (4.6). This completes the proof of the lower bound (2.4) and the theorem.  $\square$

**Lemma 4.1.** *There is a constant  $C > 0$  depending only on  $(\rho_0, \delta_0)$  such that*

$$\int_{-c_N}^0 \psi_N(s) ds \leq C N^{-\alpha_0} (\log(N))^{\frac{1}{2}(\beta_0-1)}. \tag{4.11}$$

*Proof.* To simplify the analysis of  $\psi_N(s)$ , we will introduce a change of variable and write  $s$  as function of a number  $\delta$  lying in an interval denoted as

$$\mathcal{I}_N = \left[0, 1 - \delta_0 - \frac{\frac{1}{4}\ell_2(N)}{\log(N)}\right].$$

Specifically, we write  $s = s_N(\delta) = -a_N\delta$ , where  $a_N = \sqrt{\frac{2(1-\rho_0)\log(N)}{\rho_0}}$ . Also note that  $s_N(\cdot)$  maps  $\mathcal{I}_N$  to the interval  $[-c_N, 0]$ . As a means of simplifying the factor  $\Phi^N\left(\frac{t_N - \sqrt{\rho_0}s}{\sqrt{1-\rho_0}}\right)$  in the definition of  $\psi_N(s)$ , let

$$u_N(\delta) = \frac{t_N - \sqrt{\rho_0}s_N(\delta)}{\sqrt{1-\rho_0}} = \sqrt{2\log(N)}(\delta_0 + \delta). \quad (4.12)$$

It follows that for all  $\delta \in \mathcal{I}_N$ ,

$$\log \psi_N(s_N(\delta)) = -\frac{(1-\rho_0)\log(N)\delta^2}{\rho_0} + N \log \Phi(u_N(\delta)) - \log(\sqrt{2\pi}). \quad (4.13)$$

Due to the lower-bound form of Mill's inequality, we have

$$\begin{aligned} \Phi(u_N(\delta)) &\leq 1 - \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{u_N(\delta)} - \frac{1}{u_N^3(\delta)} \right\} \exp\left(-\frac{1}{2}u_N(\delta)^2\right) \\ &= 1 - w_N(\delta), \end{aligned} \quad (4.14)$$

where  $w_N(\delta)$  is defined by the last line. When  $N$  is sufficiently large, it is simple to check that the condition  $0 < w_N(\delta) < 1$  holds for all  $\delta \in \mathcal{I}_N$ , which gives  $\log(1 - w_N(\delta)) \leq -w_N(\delta)$ . Combining the last few steps, the following bound holds for all  $\delta \in \mathcal{I}_N$ ,

$$\begin{aligned} N \log \Phi(u_N(\delta)) &\leq -N w_N(\delta) \\ &\leq \frac{-cN}{\sqrt{\log(N)}} \exp\left\{-\log(N)(\delta_0 + \delta)^2\right\}. \end{aligned} \quad (4.15)$$

The work up to this point provides us with a useful majorant for  $\psi_N$ . By looking at the equation (4.13) and the bound (4.15), it is clear that if we define the function

$$f_N(\delta) = \frac{(1-\rho_0)\log(N)\delta^2}{\rho_0} + \frac{cN}{\sqrt{\log(N)}} \exp\left\{-\log(N)(\delta_0 + \delta)^2\right\}, \quad (4.16)$$

then the bound

$$\psi_N(s_N(\delta)) \leq \exp\{-f_N(\delta)\} \quad (4.17)$$

holds for all  $\delta \in \mathcal{I}_N$ . Integrating this bound gives

$$\int_{-c_N}^0 \psi_N(s) ds \leq a_N \int_{\mathcal{I}_N} \exp\{-f_N(\delta)\} d\delta. \quad (4.18)$$

We now introduce another change of variable, and write  $\delta$  as a function of a positive number  $\eta$  using

$$\delta = \delta_N(\eta) = 1 - \delta_0 - \frac{\eta\ell_2(N)}{\log(N)}.$$

If we define the interval  $\mathcal{J}_N = [\frac{1}{4}, \frac{(1-\delta_0)\log(N)}{\ell_2(N)}]$ , then  $\delta_N(\cdot)$  maps  $\mathcal{J}_N$  to  $\mathcal{I}_N$ , and the integral bound (4.18) becomes

$$\int_{-c_N}^0 \psi_N(s) ds \leq \frac{a_N \ell_2(N)}{\log(N)} \int_{\mathcal{J}_N} \exp\{-f_N(\delta_N(\eta))\} d\eta. \quad (4.19)$$

The remainder of the proof will be divided into two parts, in which the integral over  $\mathcal{J}_N$  is decomposed with the subintervals

$$\mathcal{J}'_N = \left[\frac{1}{4}, \frac{(\log(N))^{1/4}}{\ell_2(N)}\right] \quad \text{and} \quad \mathcal{J}''_N = \left[\frac{(\log(N))^{1/4}}{\ell_2(N)}, \frac{(1-\delta_0)\log(N)}{\ell_2(N)}\right].$$

In handling these subintervals below, it will be convenient to label the summands of  $f_N$  in line (4.16) according to

$$f_N(\delta) = g_N(\delta) + h_N(\delta),$$

where

$$\begin{aligned} g_N(\delta) &= \frac{(1-\rho_0)\log(N)\delta^2}{\rho_0} \\ h_N(\delta) &= \frac{cN}{\sqrt{\log(N)}} \exp\{-\log(N)(\delta_0 + \delta)^2\}. \end{aligned}$$

**The integral over  $\mathcal{J}'_N$ .** By expanding out the square  $\delta_N(\eta)^2$ , and dropping the smallest positive term, the following bound holds for any  $\eta$ ,

$$\exp\{-g_N(\delta_N(\eta))\} \leq N^{-\alpha_0} \exp\{2\beta_0 \ell_2(N)\eta\}.$$

In addition, if we expand the square  $(\delta_0 + \delta_N(\eta))^2$ , and use the fact that every  $\eta \in \mathcal{J}'_N$  is bounded above by  $(\log(N))^{1/4}/\ell_2(N)$ , then a short calculation gives

$$-h_N(\delta_N(\eta)) \leq -ce^{\ell_2(N)(2\eta - \frac{1}{2})},$$

for small enough  $c > 0$ . Directly combining the last two steps gives

$$\exp\{-f_N(\delta_N(\eta))\} \leq N^{-\alpha_0} \exp\{2\beta_0 \eta \ell_2(N) - ce^{\ell_2(N)(2\eta - \frac{1}{2})}\}. \quad (4.20)$$

To simplify the previous bound, define  $x(\eta) = 2\eta - \frac{1}{2}$ . Since  $x(\eta)$  is non-negative for all  $\eta \in \mathcal{J}'_N$ , we may approximate  $\exp\{\ell_2(N)x(\eta)\}$  from below using a second-order Taylor expansion  $1 + \ell_2(N)x(\eta) + \frac{1}{2}\ell_2(N)^2x(\eta)^2$ . After some arithmetic, the bound (4.20) becomes

$$\exp\{-f_N(\delta_N(\eta))\} \leq e^{\frac{\beta_0}{2}\ell_2(N)} \cdot e^{-c} \cdot N^{-\alpha_0} \cdot \varphi_N(x(\eta)), \quad (4.21)$$

where we define the function

$$\varphi_N(x) = \exp\left\{-\left[\frac{c}{2}\ell_2(N)^2\right]x^2 + [(\beta_0 - c)\ell_2(N)]x\right\}.$$



Integrating the bound (4.21) over  $\mathcal{J}'_N$ , we obtain

$$\frac{a_N \ell_2(N)}{\log(N)} \int_{\mathcal{J}'_N} \exp\{-f_N(\delta_N(\eta))\} d\eta \leq C \cdot \ell_2(N) \cdot (\log(N))^{\frac{1}{2}(\beta_0-1)} \cdot N^{-\alpha_0} \int_0^\infty \varphi_N(x) dx. \quad (4.22)$$

To handle the last integral, note that the function  $\varphi_N(x)$  can be written in the form  $\varphi_N(x) = \exp(-ax^2 + bx)$ , and that the elementary Gaussian integral bound

$$\int_0^\infty e^{-ax^2+bx} dx \leq \frac{\sqrt{\pi}}{\sqrt{a}} \exp\left(\frac{b^2}{4a}\right)$$

holds for any  $a > 0$  and  $b \in \mathbb{R}$ . Therefore,

$$\int_0^\infty \varphi_N(x) dx \leq \frac{C}{\ell_2(N)}.$$

Combining this with the bound (4.22) completes the work on  $\mathcal{J}'_N$ .

**The integral over  $\mathcal{J}''_N$ .** First note that  $\exp\{-f_N(\delta)\} \leq \exp\{-h_N(\delta)\}$ . Also, the function  $\exp\{-h_N(\delta_N(\eta))\}$  is decreasing in  $\eta$ , and so if we denote the left endpoint of  $\mathcal{J}''_N$  as

$$\eta_N = (\log(N))^{1/4} / \ell_2(N),$$

then we have the following height-width integral bound

$$\begin{aligned} \frac{a_N \ell_2(N)}{\log(N)} \int_{\mathcal{J}''_N} \exp\{-f_N(\delta_N(\eta))\} d\eta &\leq \frac{a_N \ell_2(N)}{\log(N)} \cdot |\mathcal{J}''_N| \cdot \exp\{-h_N(\delta_N(\eta_N))\} \\ &\leq C \cdot \sqrt{\log(N)} \cdot \exp\left\{\frac{-c}{\sqrt{\log(N)}} e^{2(\log(N))^{1/4}}\right\}. \end{aligned}$$

This bound is of smaller order than  $N^{-\alpha_0} (\log(N))^{\frac{1}{2}(\beta_0-1)}$ , which completes the proof of Lemma 4.1.  $\square$

**Lemma 4.2.** *There is a constant  $C > 0$  depending only on  $(\delta_0, \rho_0)$  such that*

$$\int_0^\infty \psi_N(s) ds \leq C N^{-\alpha_0} (\log(N))^{\frac{1}{2}(\beta_0-1)}. \quad (4.23)$$

*Proof.* The function  $\Phi^N\left(\frac{t_N - \sqrt{\rho_0}s}{\sqrt{1-\rho_0}}\right)$  is decreasing in  $s$ , and so

$$\int_0^\infty \psi_N(s) ds \leq \Phi^N\left(\frac{t_N}{\sqrt{1-\rho_0}}\right).$$

Then, the inequality (4.15) gives

$$\begin{aligned} \Phi^N\left(\frac{t_N}{\sqrt{1-\rho_0}}\right) &\leq \exp(-h_N(0)) \\ &= \exp\left\{\frac{-cN^{1-\delta_0^2}}{\sqrt{\log(N)}}\right\}, \end{aligned} \quad (4.24)$$

which is clearly of smaller order than the stated bound.  $\square$

**Proof of Proposition 2.1.** We first show (2.1) in the case where  $\mu_N = \mathbb{E}(M_N(\xi))$ . Let  $R' \in \mathbb{R}^{N \times N}$  be a correlation matrix with  $R'_{ij} = \rho_0$  for all  $i \neq j$ , and let  $\xi' \sim \mathcal{N}(0, R')$ . Based on the assumed condition (1.1), it follows from Slepian's lemma that

$$\mathbb{E}(M_N(\xi)) \geq \mathbb{E}(M_N(\xi')). \quad (4.25)$$

Next, the representation (4.2) gives

$$\begin{aligned} \mathbb{E}[M_N(\xi')] &= \sqrt{1 - \rho_0} \mathbb{E} \left[ \max_{1 \leq j \leq N} \zeta_j \right] \\ &\geq \sqrt{2(1 - \rho_0) \log(N)} - c_0 \sqrt{\ell_2(N)} \end{aligned} \quad (4.26)$$

for some universal constant  $c_0 > 0$ , where the second step can be obtained from [Mas07, p.66]. To handle  $\mu_N = \text{med}(M_N(\xi))$ , note that

$$\begin{aligned} \left| \mathbb{E}(M_N(\xi)) - \text{med}(M_N(\xi)) \right| &\leq \sqrt{\text{var}(M_N(\xi))} \\ &\leq 1, \end{aligned} \quad (4.27)$$

where the second step uses the Gaussian Poincaré inequality [BLM13, Theorem 3.20]. To show the upper bound (2.2), note that we may take  $\xi = \xi'$ . When  $\mu_N = \mathbb{E}(M_N(\xi'))$ , the result follows from the first part of (4.26) and the standard inequality  $\mathbb{E}(\max_{1 \leq j \leq N} \zeta_j) \leq \sqrt{2 \log(N)}$ . When  $\mu_N = \text{med}(M_N(\xi'))$ , the result follows from the fact that  $\text{med}(M_N(\xi')) \leq \mathbb{E}(M_N(\xi'))$ , since  $M_N(\cdot)$  is a continuous convex function [Kwa94].  $\square$

## 5. Proof of Theorem 3.1

Let  $\Sigma = \mathbb{E}(X_1 X_1^\top)$  be the population covariance matrix, and let  $\widehat{\Sigma}$  be the sample covariance matrix defined in Section 3, with  $\widehat{\sigma}_j^2 = \widehat{\Sigma}_{jj}$  for every  $j = 1, \dots, p$ . Recall that  $S_n^*$  denotes a Gaussian random vector drawn from  $\mathcal{N}(0, \widehat{\Sigma})$ , and define a corresponding Gaussian vector  $\tilde{S}_n \sim \mathcal{N}(0, \Sigma)$ . For each index  $d \in \{1, \dots, p\}$ , let the index set  $J(d)$  be as defined in Section 3 and define three associated max statistics

$$\begin{aligned} M_d &= \max_{j \in J(d)} S_{n,j} / \sigma_j^\tau \\ \tilde{M}_d &= \max_{j \in J(d)} \tilde{S}_{n,j} / \sigma_j^\tau \\ M_d^* &= \max_{j \in J(d)} S_{n,j}^* / \widehat{\sigma}_j^\tau. \end{aligned}$$

To compare these statistics, we will use the Kolmogorov metric, defined for generic random variables  $U$  and  $V$  as

$$d_K(\mathcal{L}(U), \mathcal{L}(V)) = \sup_{t \in \mathbb{R}} |\mathbb{P}(U \leq t) - \mathbb{P}(V \leq t)|.$$

In this notation, the proof amounts to showing there is a constant  $C > 0$  not depending on  $n$  such that the event

$$d_K(\mathcal{L}(M_p), \mathcal{L}(M_p^*|X)) \leq Cn^{-\frac{1}{2}+\delta}$$

holds with probability at least  $1 - C/n$ . The overall structure of the proof will be decomposed into two parts by considering the triangle inequality

$$d_K(\mathcal{L}(M_p), \mathcal{L}(M_p^*|X)) \leq d_K(\mathcal{L}(M_p), \mathcal{L}(\tilde{M}_p)) + d_K(\mathcal{L}(\tilde{M}_p), \mathcal{L}(M_p^*|X))$$

and then separately bounding the two distances on the right. Hence, it suffices to prove the following proposition.

**Proposition 5.1.** *Suppose the conditions of Theorem 3.1 hold. Then,*

$$d_K(\mathcal{L}(M_p), \mathcal{L}(\tilde{M}_p)) \lesssim n^{-\frac{1}{2}+\delta}, \quad (i)$$

and there is a constant  $C > 0$  not depending on  $n$ , such that the event

$$d_K(\mathcal{L}(\tilde{M}_p), \mathcal{L}(M_p^*|X)) \leq Cn^{-\frac{1}{2}+\delta} \quad (ii)$$

holds with probability at least  $1 - C/n$ .

The next two subsections will give the proofs of (i) and (ii) respectively.

### 5.1. Proof of Proposition 5.1(i)

Recall  $l_n = \lceil n^{\delta/\kappa} \wedge p \rceil$  in Theorem 3.1, and define the integer

$$k_n = l_n^2 \wedge p.$$

We proceed by considering the bound

$$d_K(\mathcal{L}(M_p), \mathcal{L}(\tilde{M}_p)) \leq \text{I}_n + \text{II}_n + \text{III}_n,$$

where the terms on the right are defined by

$$\text{I}_n = d_K(\mathcal{L}(M_p), \mathcal{L}(M_{k_n})) \quad (5.1)$$

$$\text{II}_n = d_K(\mathcal{L}(M_{k_n}), \mathcal{L}(\tilde{M}_{k_n})) \quad (5.2)$$

$$\text{III}_n = d_K(\mathcal{L}(\tilde{M}_{k_n}), \mathcal{L}(\tilde{M}_p)). \quad (5.3)$$

In the remainder of this subsection, we will focus on establishing the bound

$$\text{I}_n \lesssim n^{-\frac{1}{2}+\delta}. \quad (5.4)$$

This will be sufficient to prove Proposition 5.1(i), because the quantity  $\text{III}_n$  can be bounded using the same argument as for  $\text{I}_n$ , and because the bound

$$\text{II}_n \lesssim n^{-\frac{1}{2}+\delta} \quad (5.5)$$

follows from Lemma 5.1 of Section 5.3.

Going forward, including in Section 5.2, we may assume without loss of generality that  $k_n < p$ , because if  $p \leq k_n$ , then  $p = k_n$  and the quantities  $I_n$  and  $III_n$  become identically 0. Furthermore, the argument for bounding  $II_n$  does not depend on the relationship between  $k_n$  and  $p$ . To proceed, first notice that the definition of  $I_n$  gives

$$\begin{aligned} I_n &= \sup_{t \in \mathbb{R}} \left| \mathbb{P} \left( \max_{1 \leq j \leq p} S_{n,j} / \sigma_j^\tau \leq t \right) - \mathbb{P} \left( \max_{j \in J(k_n)} S_{n,j} / \sigma_j^\tau \leq t \right) \right| \\ &= \sup_{t \in \mathbb{R}} \mathbb{P}(A(t) \cap B(t)), \end{aligned}$$

where we define the following events for arbitrary  $t \in \mathbb{R}$ ,

$$A(t) = \left\{ \max_{j \in J(k_n)} S_{n,j} / \sigma_j^\tau \leq t \right\} \quad \text{and} \quad B(t) = \left\{ \max_{j \in J(k_n)^c} S_{n,j} / \sigma_j^\tau > t \right\}.$$

For any real numbers  $t_{1,n}$  and  $t_{2,n}$  satisfying  $t_{1,n} \leq t_{2,n}$ , it can be checked that the following inclusion holds simultaneously for all  $t \in \mathbb{R}$ ,

$$A(t) \cap B(t) \subset A(t_{2,n}) \cup B(t_{1,n}).$$

Therefore, a union bound gives

$$I_n \leq \mathbb{P}(A(t_{2,n})) + \mathbb{P}(B(t_{1,n})). \quad (5.6)$$

In order to choose suitable values of  $t_{1,n}$  and  $t_{2,n}$ , define the parameter

$$\omega = 2\delta^2 / \kappa,$$

which satisfies  $\omega \in (0, \delta)$  for any choice of  $\delta \in (0, 1/2)$ . Also, define the integer

$$d_n = \lfloor (\frac{\omega^2}{4} \mathbf{r}(R(l_n))) \vee 1 \rfloor.$$

Note that since  $\mathbf{r}(R(l_n)) \leq l_n$ , it is clear that the inequalities  $d_n \leq l_n \leq k_n \leq p$  hold for all  $n$ . Also, using the assumptions of Theorem 3.1, we have the following lower bound on  $d_n$ ,

$$d_n \gtrsim \mathbf{r}(R(l_n)) = \frac{l_n^2}{\|R(l_n)\|_F^2} \gtrsim l_n^\delta \gtrsim n^{\delta^2/\kappa}, \quad (5.7)$$

which will be used later.

To finish the proof of (5.4), it suffices to exhibit values  $t_{1,n}$  and  $t_{2,n}$  that satisfy the following three conditions:

$$t_{1,n} \leq t_{2,n} \quad \text{holds for all large } n \quad (5.8)$$

$$\mathbb{P}(B(t_{1,n})) \lesssim n^{-1} \quad (5.9)$$

$$\mathbb{P}(A(t_{2,n})) \lesssim n^{-1/2+\delta}. \quad (5.10)$$

In order to specify such values of  $t_{1,n}$  and  $t_{2,n}$ , we will take them to be of the form

$$\begin{aligned} t_{1,n} &= c_1 \cdot k_n^{-\gamma(1-\tau)} \cdot \log(n) \\ t_{2,n} &= c_2 \cdot l_n^{-\gamma(1-\tau)} \cdot \sqrt{\log(d_n)}, \end{aligned}$$

for certain constants  $c_1, c_2 > 0$  that do not depend on  $n$ . First, the condition (5.8) holds for any fixed choices of  $c_1$  and  $c_2$ , due to the definitions of  $k_n$ ,  $l_n$ , and  $d_n$ .

Second, to establish (5.9), define the parameter  $q = \max\{2/(\gamma(1-\tau)), \log(n), 3\}$ . For any  $t > 0$ , we have the tail bound

$$\mathbb{P}(B(t)) \leq t^{-q} \left\| \max_{j \in J(k_n)^c} S_{n,j}/\sigma_j^\tau \right\|_q^q, \quad (5.11)$$

where  $\|\cdot\|_q$  denotes the usual  $L_q$  norm of a random variable. Due to Lemma 5.2 in Section 5.3, we have  $\|\frac{1}{\sigma_j} S_{n,j}\|_q \leq cq$  for all  $j = 1, \dots, p$ , and so

$$\begin{aligned} \left\| \max_{j \in J(k_n)^c} S_{n,j}/\sigma_j^\tau \right\|_q^q &\leq \sum_{j \in J(k_n)^c} \|S_{n,j}/\sigma_j^\tau\|_q^q \\ &\leq (cq)^q \sum_{j \in J(k_n)^c} \sigma_j^{q(1-\tau)} \\ &\lesssim (cq)^q \sum_{j=k_n+1}^p j^{-q\gamma(1-\tau)} \\ &\lesssim \frac{(cq)^q}{q\gamma(1-\tau)-1} k_n^{-q\gamma(1-\tau)+1}. \end{aligned}$$

One can check that  $q \asymp \log(n)$  and  $\frac{c}{(q\gamma(1-\tau)-1)^{1/q}} k_n^{1/q} \lesssim 1$ . Therefore, if we take

$$t = e \cdot \left\| \max_{j \in J(k_n)^c} S_{n,j}/\sigma_j^\tau \right\|_q$$

in (5.11), then there is a choice of  $c_1$  for which  $t_{1,n}$  satisfies

$$t \leq t_{1,n},$$

and furthermore,

$$\mathbb{P}(B(t_{1,n})) \leq \mathbb{P}(B(t)) \leq e^{-q} \leq n^{-1},$$

as needed for (5.9).

Third, we now turn to (5.10). Observe that

$$\begin{aligned} \mathbb{P}(A(t_{2,n})) &\leq \mathbb{P}\left(\max_{j \in J(k_n)} \tilde{S}_{n,j}/\sigma_j^\tau \leq t_{2,n}\right) + \text{II}_n \\ &\lesssim \mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j}/\sigma_j^\tau \leq t_{2,n}\right) + n^{-1/2+\delta} \end{aligned} \quad (5.12)$$

where the second step follows from (5.5) and the inclusion  $J(l_n) \subset J(k_n)$ .

As a preparatory step towards applying Theorem 2.3 to the last line of (5.12), we need the following basic observation. Let  $(Y_j)_{j \in J(l_n)}$  be a generic set of random variables, and let  $(a_j)_{j \in J(l_n)}$  be positive numbers satisfying the condition  $\max_{j \in J(l_n)} a_j \leq b$ , for some number  $b$ . It is straightforward to check that

$$\mathbb{P}\left(\max_{j \in J(l_n)} Y_j \leq t_{2,n}\right) \leq \mathbb{P}\left(\max_{j \in J(l_n)} a_j Y_j \leq b t_{2,n}\right). \quad (5.13)$$

Based on Assumption 3.1, there is a positive constant  $c_0$  not depending on  $n$  such that the inequality  $\sigma_j^{-(1-\tau)} \leq l_n^{\gamma(1-\tau)}/c_0$  holds for all  $j \in J(l_n)$ . Accordingly, we will use (5.13) with the choices  $a_j = \sigma_j^{-(1-\tau)}$ ,  $b = l_n^{\gamma(1-\tau)}/c_0$ , and  $Y_j = \tilde{S}_{n,j}/\sigma_j^\tau$ . Also, we may choose  $c_2$  in the definition of  $t_{2,n}$  to have the value  $c_2 = \omega c_0 \sqrt{2(1-\omega)}$ , which implies  $b t_{2,n} = \omega \sqrt{2(1-\omega) \log(d_n)}$ . Under these choices, the inequality (5.13) becomes

$$\mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j}/\sigma_j^\tau \leq t_{2,n}\right) \leq \mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j}/\sigma_j \leq \omega \sqrt{2(1-\omega) \log(d_n)}\right).$$

Now, we apply Theorem 2.3 to the right side, with  $(l_n, d_n, \omega, \omega)$  playing the roles of  $(N, k, \epsilon, \delta)$  in that statement of that result. (Under these choices, the application of Theorem 2.3 is justified because  $\omega \in (0, 1)$  and the inequalities  $d_n \leq (\omega^2/4)\mathfrak{r}(R(l_n))$  and  $d_n \geq 2$  hold for all large  $n$ .) Hence,

$$\mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j}/\sigma_j^\tau \leq t_{2,n}\right) \lesssim d_n^{\frac{-(1-\omega)(1-\omega)^2}{\omega}} (\log(d_n))^{\frac{1-\omega(2-\omega)-\omega}{2\omega}}.$$

Furthermore, using the lower bound on  $d_n$  from (5.7), there is a constant  $c$  not depending on  $n$  such that

$$\begin{aligned} \mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j}/\sigma_j^\tau \leq t_{2,n}\right) &\lesssim (n^{\delta^2/\kappa})^{\frac{-(1-\omega)^3}{\omega}} \cdot (\log(n))^c \\ &= n^{-\frac{(1-\omega)^3}{2}} \cdot (\log(n))^c \\ &\lesssim n^{-\frac{1}{2}+3\omega} \\ &\lesssim n^{-\frac{1}{2}+\delta} \end{aligned}$$

which completes the proof.  $\square$

## 5.2. Proof of Proposition 5.1(ii)

The proof of part (ii) is structured mostly along the same lines as the proof of part (i), and so we only sketch out the main steps for the sake of brevity. The current proof will also continue to use the same notation. Consider the inequality

$$d_K(\mathcal{L}(\tilde{M}_p), \mathcal{L}(M_p^*|X)) \leq \text{I}'_n + \text{II}'_n(X) + \text{III}'_n(X),$$

where we define

$$I'_n = d_K(\mathcal{L}(\tilde{M}_p), \mathcal{L}(\tilde{M}_{k_n})) \quad (5.14)$$

$$II'_n(X) = d_K(\mathcal{L}(\tilde{M}_{k_n}), \mathcal{L}(M_{k_n}^*|X)) \quad (5.15)$$

$$III'_n(X) = d_K(\mathcal{L}(M_{k_n}^*|X), \mathcal{L}(M_p^*|X)). \quad (5.16)$$

Note that  $I'_n$  is non-random, whereas  $II'_n(X)$  and  $III'_n(X)$  are random.

Establishing a bound on  $I'_n$  of order  $n^{-1/2+\delta}$  requires no further work, because  $I'_n$  is equal to  $III_n$  in the proof of part (i). Next, it follows from Lemma 5.1 in Section 5.3 that the event

$$II'_n(X) \leq C n^{-\frac{1}{2}+\delta} \quad (5.17)$$

holds with probability at least  $1 - C/n$ . So, it remains to establish a bound on  $III'_n(X)$  of order  $n^{-1/2+\delta}$ . For this purpose, let  $t'_{1,n}$  and  $t'_{2,n}$  be any real numbers satisfying  $t'_{1,n} \leq t'_{2,n}$ . The reasoning leading up to (5.6) can be re-used to show that the following bound holds almost surely

$$III'_n(X) \leq \mathbb{P}(A'(t'_{2,n})|X) + \mathbb{P}(B'(t'_{1,n})|X), \quad (5.18)$$

where we define the following events for arbitrary  $t \in \mathbb{R}$ ,

$$A'(t) = \left\{ \max_{j \in J(k_n)} S_{n,j}^* / \hat{\sigma}_j^\tau \leq t \right\} \quad \text{and} \quad B'(t) = \left\{ \max_{j \in J(k_n)^c} S_{n,j}^* / \hat{\sigma}_j^\tau > t \right\}.$$

To bound the two probabilities on the right side of (5.18), we will use values  $t'_{1,n}$  and  $t'_{2,n}$  having the form

$$t'_{1,n} = c'_1 \cdot k_n^{-\gamma(1-\tau)} \cdot (\log(n))^{3/2}$$

$$t'_{2,n} = c'_2 \cdot l_n^{-\gamma(1-\tau)} \cdot \sqrt{\log(d_n)},$$

for certain constants  $c'_1, c'_2 > 0$  that do not depend on  $n$ . (Note that for any such choices, the condition  $t'_{1,n} \leq t'_{2,n}$  will hold for all large  $n$ .)

The probability  $\mathbb{P}(A'(t'_{2,n})|X)$  can be handled using the argument after (5.12) in the proof of part (i), together with the bound (5.17). Specifically, it can be shown that there are constants  $C$  and  $c'_2$  such that

$$\begin{aligned} \mathbb{P}(A'(t'_{2,n})|X) &\leq \mathbb{P}\left(\max_{j \in J(l_n)} \tilde{S}_{n,j} / \sigma_j^\tau \leq t'_{2,n}\right) + II'_n(X) \\ &\leq C n^{-\frac{1}{2}+\delta} \end{aligned}$$

holds with probability at least  $1 - C/n$ . Finally, an argument analogous to the one used to establish (5.9) earlier shows there are constants  $C$  and  $c'_1$  such that

$$\mathbb{P}(B'(t'_{1,n})|X) \leq n^{-1}$$

holds with probability at least  $1 - C/n$ . (A more detailed version of this argument can be found in the proof of Lemma C.1 part (b) in [LLM20].)  $\square$

### 5.3. Background results

The two bounds (5.19) and (5.20) below follow from the proofs of Propositions B.1 and C.1 in [LLM20]. These proofs can be applied in the setting of Theorem (3.1) in this paper with no essential changes.

**Lemma 5.1.** *Fix any  $\delta \in (0, 1/2)$ , and suppose the conditions of Theorem 3.1 hold. Also, let  $\Pi_n$  and  $\Pi'_n(X)$  be as defined in (5.2) and (5.15) respectively. Then,*

$$\Pi_n \lesssim n^{\frac{1}{2}+\delta}, \quad (5.19)$$

and there is a constant  $C > 0$  not depending on  $n$  such that the event

$$\Pi'_n(X) \leq Cn^{-\frac{1}{2}+\delta} \quad (5.20)$$

holds with probability at least  $1 - C/n$ .

For the next lemma, recall that  $\|\cdot\|_q$  denotes the  $L_q$  norm of a random variable. This lemma is effectively a restatement of Lemma D.4 in [LLM20], and the proof given there can be applied in the same manner under the conditions used here.

**Lemma 5.2.** *Suppose the conditions of Theorem 3.1 hold, and define the parameter  $q = \max\{2/(\gamma(1-\tau)), \log(n), 3\}$ . Then,*

$$\max_{1 \leq j \leq p} \left\| \frac{1}{\sigma_j} S_{n,j} \right\|_q \lesssim q.$$

## References

- [Adl90] R. J. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics, 1990. [MR1088478](#)
- [BL08] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Annals of Statistics*, 36(1):199–227, 2008. [MR2387969](#)
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. [MR3185193](#)
- [Bor75] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975. [MR0399402](#)
- [BT87] J. Bourgain and L. Tzafriri. Invertibility of large submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel Journal of Mathematics*, 57(2):137–224, 1987. [MR0890420](#)
- [CCK13] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics*, 41(6):2786–2819, 2013. [MR3161448](#)
- [CCK20] V. Chernozhukov, D. Chetverikov, and Y. Koike. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *arXiv:2012.09513*, 2020.



- [CCKK19] V. Chernozhukov, D. Chetverikov, K. Kato, and Y. Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *arXiv:1912.10529*, 2019. [MR3693963](#)
- [Cha14] S. Chatterjee. *Superconcentration and Related Topics*. Springer, 2014. [MR3157205](#)
- [DEZ15] J. Ding, R. Eldan, and A. Zhai. On multiple peaks and moderate deviations for the supremum of a Gaussian field. *Annals of Probability*, 43(6):3468–3493, 2015. [MR3433585](#)
- [Dur05] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, 3rd edition, 2005. [MR1609153](#)
- [DZ20] H. Deng and C.-H. Zhang. Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Annals of Statistics*, 48(6):3643–3671, 2020. [MR4185823](#)
- [GR15] D. Guillet and B. Rajaratnam. Functions preserving positive definiteness for sparse matrices. *Transactions of the American Mathematical Society*, 367(1):627–649, 2015. [MR3271272](#)
- [Har14] J. A. Hartigan. Bounding the maximum of dependent random variables. *Electronic Journal of Statistics*, 8(2):3126–3140, 2014. [MR3301303](#)
- [Kim19] K. Kim. On the large-scale structure of the tall peaks for stochastic heat equations with fractional Laplacian. *Stochastic Processes and their Applications*, 129(6):2207–2227, 2019. [MR3958428](#)
- [KKX17] D. Khoshnevisan, K. Kim, and Y. Xiao. Intermittency and multifractality: A case study via parabolic stochastic PDEs. *Annals of Probability*, 45(6A):3697–3751, 2017. [MR3729613](#)
- [KKX18] D. Khoshnevisan, K. Kim, and Y. Xiao. A macroscopic multifractal analysis of parabolic stochastic PDEs. *Communications in Mathematical Physics*, 360(1):307–346, 2018. [MR3795193](#)
- [Kwa94] S. Kwapień. A remark on the median and the expectation of convex functions of Gaussian vectors. In *Probability in Banach Spaces, 9*, pages 271–272. Springer, 1994. [MR1308523](#)
- [Lif95] M. A. Lifshits. *Gaussian Random Functions*. Springer, 1995. [MR1472736](#)
- [LLM20] M. E. Lopes, Z. Lin, and H.-G. Müller. Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Annals of Statistics*, 48(2):1214–1229, 2020. [MR4102694](#)
- [LLR83] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983. [MR0691492](#)
- [Lop20] M. E. Lopes. Central limit theorem and bootstrap approximation in high dimensions with near  $1/\sqrt{n}$  rates. *arXiv:2009.06004*, 2020.
- [LS01] W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. *Handbook of Statistics*, 19:533–597, 2001. [MR1861734](#)
- [LT13] M. Ledoux and M. Talagrand. *Probability in Banach Spaces:*

- Isoperimetry and Processes*. Springer, 2013. [MR2814399](#)
- [Mas07] P. Massart. *Concentration Inequalities and Model Selection*. Springer, 2007. [MR2319879](#)
- [MSS15] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. *Annals of Mathematics*, pages 327–350, 2015. [MR3374963](#)
- [MSS21] A. W. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families III: Sharper restricted invertibility estimates. *Israel Journal of Mathematics (to appear)*, 2021. [MR3374962](#)
- [Pól49] George Pólya. Remarks on characteristic functions. In *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, pages 115–123. University of California Press, 1949. [MR0028541](#)
- [PV18] G. Paouris and P. Valettas. A Gaussian small deviation inequality for convex functions. *Annals of Probability*, 46(3):1441–1454, 2018. [MR3785592](#)
- [PV19] G. Paouris and P. Valettas. Variance estimates and almost Euclidean structure. *Advances in Geometry*, 2019. [MR3940918](#)
- [Sch07] G. Schechtman. The random version of Dvoretzky’s theorem in  $\ell_\infty^n$ . In *Geometric Aspects of Functional Analysis*, pages 265–270. 2007. [MR2349612](#)
- [Sle62] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962. [MR0133183](#)
- [ST74] V. N. Sudakov and B. S. Tsirel’son. Extremal properties of semi-spaces for spherically symmetrical measures (in Russian). *Zapiski Nauchnykh Seminarov POMI*, 41:14–24, 1974. [MR0365680](#)
- [Tal14] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer, 2014. [MR3184689](#)
- [Tan15] K. Tanguy. Some superconcentration inequalities for extrema of stationary Gaussian processes. *Statistics & Probability Letters*, 106:239–246, 2015. [MR3389997](#)
- [Tan19] K. Tanguy. Improved one-sided deviation inequalities under regularity assumptions for product measures. *ESAIM: Probability and Statistics*, 23:979–990, 2019. [MR4046859](#)
- [Val19] P. Valettas. On the tightness of Gaussian concentration for convex functions. *Journal d’Analyse Mathématique*, 2019. [MR4041105](#)
- [Yi21] J. Yi. Macroscopic multi-fractality of Gaussian random fields and linear SPDEs with colored noise. *arXiv:2101.09450*, 2021.