On the rate of concentration of maxima in Gaussian arrays

Rafail Kartsioukas[1] · Zheng Gao[1] · Stilian Stoev[1] (ORCID)

Abstract

Recently in Gao and Stoev (2020) it was established that the concentration of maxima phenomenon is the key to solving the exact sparse support recovery problem in high dimensions. This phenomenon, known also as relative stability, has been little studied in the context of dependence. Here, we obtain bounds on the rate of concentration of maxima in Gaussian triangular arrays. These results are used to establish sufficient conditions for the uniform relative stability of functions of Gaussian arrays, leading to new models that exhibit phase transitions in the exact support recovery problem. Finally, the optimal rate of concentration for Gaussian arrays is studied under general assumptions implied by the classic condition of Berman (1964).

Keywords Rate of relative stability · Concentration of maxima ·
Exact support recovery · Phase transitions · Functions of Gaussian arrays

AMS 2000 Subject Classifications MSC 62G32 · 62G20 · 62G10 · 60G15 · 60G70

1 Introduction

Let Z_i, $i = 1, 2, \ldots$ be independent and identically distributed (iid) standard Normal random variables. It is well known that their maxima under affine normalization converge to the Gumbel extreme value distribution. If, however, one chooses to

✉ Stilian Stoev
  sstoev@umich.edu

  Rafail Kartsioukas
  rkarts@umich.edu

  Zheng Gao
  gaozheng@umich.edu

[1] Department of Statistics, University of Michigan, Ann Arbor, MI, USA

standardize the maxima by only dividing by a sequence of positive numbers, then the only possible limits are constants. Specifically, for all $a_p \sim \sqrt{2 \log(p)}$, we have

$$\frac{1}{a_p} \max_{i \in [p]} Z_i \xrightarrow{\mathbb{P}} 1, \quad \text{as } p \to \infty, \tag{1.1}$$

where $[p] := \{1, \cdots, p\}$ and in fact the convergence is valid almost surely. This property, known as *relative stability*, dates back to the seminal work of Gnedenko (1943) who has characterized it in terms of rapid variation of the law of the Z_i's (see Section 2.2 below, as well as Barndorff-Nielsen 1963; Resnick and Tomkins 1973; Kinoshita and Resnick 1991).

In contrast, if the Z_i's are iid and heavy-tailed, i.e., $\mathbb{P}[Z_i > x] \propto x^{-\alpha}$, for some $\alpha > 0$, with $a_p \propto p^{1/\alpha}$, we have

$$\frac{1}{a_p} \max_{i \in [p]} Z_i \xrightarrow{d} \xi, \tag{1.2}$$

where ξ is a random variable with the α-Fréchet distribution.

Comparing (1.1) and (1.2), we see that the maxima have fundamentally different asymptotic behavior relative to rescaling with constant sequences. In the light-tailed regime, they *concentrate* around a constant in the sense of (1.1), whereas in the heavy-tailed regime they *disperse* according to a probability distribution viz (1.2).

Although this concentration of maxima phenomenon may be well-known under independence, we found that it is virtually unexplored under *dependence*. In this paper, we will focus on Gaussian sequences, and in fact, more generally, Gaussian triangular arrays $\mathscr{E} = \{\epsilon_p(i), i \in [p], p \in \mathbb{N}\}$, where the $\epsilon_p(i)$'s are marginally standard Normal but possibly dependent. Let u_p be the $(1 - 1/p)$-th quantile of the standard Normal distribution, i.e., $p\overline{\Phi}(u_p) := p(1 - \Phi(u_p)) = 1$. We say that the array $\mathscr{E}$ is uniformly relatively stable (URS), if

$$\frac{1}{u_{|S_p|}} \max_{i \in S_p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \quad \text{as } |S_p| \to \infty, \tag{1.3}$$

for every choice of growing subsets $S_p \subset \{1, \cdots, p\}$. Note that $u_p \sim \sqrt{2 \log(p)}$ (see *e.g.* Lemma 3.1). Certainly, the relative stability property shows that all iid Gaussian arrays are trivially URS. The notion of uniform relative stability, however, is far from automatic or trivial under dependence. In the recent work of Gao and Stoev (2020), it was found that URS is the key to establishing the fundamental limits in sparse-signal support estimation in high-dimensions. Specifically, under URS, a phase-transition phenomenon was shown to take place in the support recovery problem. For more details, see Section 2.1 below.

Theorem 3.1 in Gao and Stoev (2020) gives a surprisingly simple necessary and sufficient condition for a Gaussian array $\mathscr{E}$ to be URS. As an illustration, in the special case where $\epsilon_p(i) \equiv Z_i$, $i \in \mathbb{N}$ form a stationary Gaussian time series, the array $\mathscr{E}$ is URS if and only if the auto-covariance vanishes, i.e.,

$$\text{Cov}(Z_k, Z_0) \longrightarrow 0, \quad \text{as } k \to \infty. \tag{1.4}$$

That is, (1.1) holds (with $a_p \sim \sqrt{2\log(p)}$), for *any* stationary Gaussian time series $Z = \{Z_i\}$ with vanishing auto-covariance, no matter the rate of decay. The "if" part of Eq. 1.4 appeared in Theorem 4.1 in Berman (1964).

Condition (1.4) should be contrasted with the classic Berman condition,

$$\text{Cov}(Z_k, Z_0) = o\left(\frac{1}{\log(k)}\right), \quad \text{as } k \to \infty,$$

which entails *distributional* convergence under affine normalization. Here, our focus is not on distributional limits but on merely the concentration of maxima under rescaling, which can take place under much more severe dependence. In fact, unlike Berman, here we are not limited to the time-series setting. For a complete statement of the characterization of URS, see Section 2.2, below.

While Gao and Stoev (2020) characterized the conditions under which the convergence (1.3) takes place, the rate of this convergence remained an open question. In this paper, our goal is to establish bounds on the *rate of concentration* for maxima of Gaussian arrays. Specifically, we establish results of the type

$$\mathbb{P}\left[\left|\frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i) - 1\right| > \delta_p\right] \longrightarrow 0, \tag{1.5}$$

where $\delta_p \to 0$ decays at a certain rate. The rate of the sequence δ_p is quantified explicitly in terms of the covariance structure of the array. More precisely, the packing numbers $N(\tau)$ associated with the UDD condition introduced in Gao and Stoev (2020) will play a key role. These packing numbers arise from a Sudakov-Fernique type construction, which appear to be close to optimal, although at this point we do not know if the so obtained bounds on the rates can be improved (cf Conjecture 1, below). After completing this work, we became aware of the important results of Tanguy (2015), which are closely related to ours in the special case of stationary time series. Our approach, however, is technically different and yields explicit rates for the general case of Gaussian triangular arrays. For more details, see Remark 5.2, below.

Our general results are illustrated with several models, where explicit bounds on the rates of concentration are derived. In Section 3, we study the *optimal rate* of concentration and show that under rather broad dependence conditions (including the iid setting), Eq. 1.5 holds if and only if $\delta_p \gg 1/\log(p)$. Somewhat curiously, the constant u_p matters and the popular choice of $u_p := \sqrt{2\log(p)}$ leads to the slower rates of $\log(\log(p))/\log(p)$.

Our bounds on the rate of concentration find important application in the study of uniform relative stability for functions of Gaussian arrays. Specifically, let $\eta_p(i) = f(\epsilon_p(i))$, where $\mathscr{E} = \{\epsilon_p(i), \ i \in [p], \ p \in \mathbb{N}\}$ is a Gaussian triangular array and f is a given deterministic function. In Section 4.2, using our results on the rate of concentration for the array $\mathscr{E}$, we establish conditions which imply the uniform relative stability of the array $\mathscr{H} = \{\eta_p(i), \ i \in [p], \ p \in \mathbb{N}\}$. Consequently, we establish that many dependent log-normal and χ^2-arrays are URS, and hence obey the phase-transition result of Gao and Stoev (2020).

The paper is structured as follows. In Section 2, we review the statistical inference problem motivating the study of the concentration of maxima phenomenon. Recalled is the notion of uniform decreasing dependence involved in the characterization of

uniform relative stability for Gaussian arrays. A brief discussion on the optimal rate of concentration is given in Section 3. Section 4 contains the statement of the main result as well as some examples and applications. Section 5 contains proofs and technical results, which may be of independent interest.

2 Concentration of maxima and high-dimensional inference

In this section, we start with the statistical inference problem that motivated us to study the concentration of maxima phenomenon. Readers who are convinced that this is a phenomenon of independent interest can skip to Section 2.2, where concrete definitions and notions are reviewed.

2.1 Fundamental limits of support recovery in high dimensions

Our main motivation to study the relative stability or *concentration of maxima* under dependence is the fundamental role it plays in recent developments on high-dimensional statistical inference, which we briefly review next. Consider the classic *signal plus noise* model

$$x_p(i) = \mu_p(i) + \epsilon_p(i), \quad i \in [p],$$

where $\mu_p = (\mu_p(i)) \in \mathbb{R}^p$ is an unknown high-dimensional 'signal' observed with additive noise. The noise is modeled with a triangular array $\mathscr{E} = \{\epsilon_p(i), \ i \in [p], \ p \in \mathbb{N}\}$, where for concreteness, all $\epsilon_p(i)$'s are standardized to have the same marginal distribution F. However, this noise can have arbitrary dependence structure, in principle.

One popular and important high-dimensional inference context, is the one where the dimension p grows to infinity and the signal is sparse. Namely, the signal support set $S_p := \{i \in [p] : \mu_p(i) \neq 0\}$ is of smaller order than its dimension:

$$|S_p| \sim p^{1-\beta}, \text{ for some } \beta \in (0, 1).$$

The parameter β controls the degree of sparsity; if β is larger, the signal is more sparse, i.e., has fewer non-zero components. In this context, many natural questions arise such as the detection of the presence of non-zero signal or the estimation of its support set (see, e.g., Ingster 1998; Donoho and Jin 2004; Ji and Jin 2012; Arias-Castro and Chen 2017). Here, as in Gao and Stoev (2020), we focus on the fundamental *support recovery* problem. Particularly, under what conditions on the signal magnitude we can have *exact support recovery* in the sense that

$$\mathbb{P}[\widehat{S}_p = S_p] \longrightarrow 1, \quad \text{as } p \to \infty.$$

Gao and Stoev (2020) showed that a natural solution to this problem can be obtained using the concentration of maxima phenomenon. Specifically, consider the class of all thresholding support estimators:

$$\widehat{S}_p := \{j \in [p] : x_p(j) > t_p(x)\}, \tag{2.1}$$

where $t_p(x)$ is possibly data-dependent threshold. For simplicity of exposition, suppose also that the signal magnitude is parametrized as follows

$$\mu_p(i) = \sqrt{2r \log(p)}, \quad i \in S_p,$$

where $r > 0$. Consider also the function

$$g(\beta) := (1 + \sqrt{1 - \beta})^2.$$

Theorems 2.1 and 2.2 of Gao and Stoev (2020) entail that if $\mathscr{E}$ is URS (see Definition 2.2 below), then we have the phase-transition:

$$\mathbb{P}[\widehat{S}_p = S_p] \longrightarrow \begin{cases} 1, & \text{if } r > g(\beta) \text{ for suitable } \widehat{S}_p \text{ as in Eq. 2.1} \\ 0, & \text{if } r < g(\beta) \text{ for all } \widehat{S}_p \text{ as in Eq. 2.1} \end{cases}, \quad \text{as } p \to \infty.$$

That is, for signal magnitudes above the boundary, thresholding (Bonferonni-type) estimators recover the support perfectly, as $p \to \infty$; whereas for signals below the boundary, no thresholding estimators can recover the support with positive probability. Further, as shown in Gao and Stoev (2020), thresholding estimators are optimal in the iid Gaussian setting and hence the above phase-transition applies to *all possible* support estimators leading to minimax-type results. Interestingly, both Gaussian and non-Gaussian noise arrays are addressed equally well, provided that they satisfy the uniform relative stability property. While URS is a very mild condition, except for the Gaussian case addressed in Gao and Stoev (2020), little is known in general. Here, we will fill this gap for a class of functions of Gaussian arrays (see Section 4.2), using our new results on the rates of concentration.

2.2 Concentration of maxima

In this section, we recall some definitions and a characterization of URS in Gao and Stoev (2020). We start by presenting the notion of *relative stability*.

Definition 2.1 (Relative stability) Let $\epsilon_p = (\epsilon_p(j))_{j=1}^p$ be a sequence of random variables with identical marginal distributions F. Define the sequence $(u_p)_{p=1}^\infty$ to be the $(1 - 1/p)$-th quantile of F, i.e.,

$$u_p = F^{\leftarrow}(1 - 1/p). \tag{2.2}$$

The triangular array $\mathscr{E} = \{\epsilon_p, \ p \in \mathbb{N}\}$ is said to have relatively stable (RS) maxima if

$$\frac{1}{u_p} M_p := \frac{1}{u_p} \max_{i=1,\dots,p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \tag{2.3}$$

as $p \to \infty$.

Note that by Proposition 1.1 of Gao and Stoev (2020), we have for the standard Normal distribution, that

$$u_p = \Phi^{\leftarrow}(1 - 1/p) \sim \sqrt{2 \log(p)}. \tag{2.4}$$

While relative stability is not directly used in this paper, it is a natural prerequisite to introducing the following generalization.

Definition 2.2 (Uniform Relative Stability (URS)) Under the notations established in Definition 2.1, the triangular array $\mathscr{E} = \{\epsilon_p(i),\ i \in [p]\}$ is said to have uniform relatively stable (URS) maxima if for *every* sequence of subsets $S_p \subseteq \{1, \ldots, p\}$ such that $|S_p| \to \infty$, we have

$$\frac{1}{u_{|S_p|}} M_{S_p} := \frac{1}{u_{|S_p|}} \max_{i \in S_p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \quad \text{as } p \to \infty. \tag{2.5}$$

Definition 2.3 (Uniformly Decreasing Dependence (UDD)) A Gaussian triangular array $\mathscr{E}$ with standard normal marginals is said to be uniformly decreasingly dependent (UDD) if for every $\tau > 0$ there exists a finite $N_{\mathscr{E}}(\tau) < \infty$, such that for every $i \in \{1, \ldots, p\}$, and $p \in \mathbb{N}$, we have

$$\left| \{k \in \{1, \ldots, p\} : \text{Cov}(\epsilon_p(k), \epsilon_p(i)) > \tau\} \right| \le N_{\mathscr{E}}(\tau), \quad \text{for all } \tau > 0. \tag{2.6}$$

That is, for any coordinate j, the number of coordinates which are more than τ-correlated with $\epsilon_p(j)$ does not exceed $N_{\mathscr{E}}(\tau)$.

The next result provides the equivalence between uniform relative stability and uniformly decreasing dependence.

Theorem 2.1 (Theorem 3.2 in Gao and Stoev (2020)) *Let $\mathscr{E}$ be a Gaussian triangular array with standard Normal marginals. The array $\mathscr{E}$ is URS if and only if it is UDD.*

Theorem 2.1 is the starting point of the rate investigations in our paper. Our main result, Theorem 4.1, below, extends the former by providing upper bounds on the rate of concentration. Before that, though, in Section 3 we study cases where the optimal rate can be formally established.

Remark 2.1 (*On the use of the term "upper bound"*) Fix a positive sequence $\delta_p^\star \downarrow 0$. We refer to $\delta_p^\star$ as an upper bound on the rate of concentration when Eq. 1.5 holds for any sequence $\delta_p \gg \delta_p^\star$. Further, for two positive sequences α_p and β_p we write $\alpha_p \asymp \beta_p$ if

$$0 < c_1 \le \liminf_{p \to \infty} \left| \frac{\alpha_p}{\beta_p} \right| \le \limsup_{p \to \infty} \left| \frac{\alpha_p}{\beta_p} \right| \le c_2 < \infty.$$

Let $\delta_p^\star$ be an upper bound on the rate of concentration and $\delta_p \gg \delta_p^\star$. Then, naturally, Eq. 1.5 holds with δ_p replaced by $\tilde{\delta}_p$, for any $\tilde{\delta}_p \asymp \delta_p$.

3 On the optimal rate of concentration

In this section, we provide some general comments on the fastest possible rates of concentration for maxima of Gaussian variables. Somewhat surprisingly, the rate depends on the choice of the normalizing sequence u_p. As it turns out poor choices of normalizing sequences can lead to arbitrarily slow rates. On the other hand, for a

wide range of dependence structures (including the iid case), the best possible rate will be shown to be $1/\log(p)$. The question of whether the maxima of dependent Gaussian arrays can concentrate faster than that rate, however unlikely this may be, is open, to the best of our knowledge (cf Conjecture 1, below).

Consider a Gaussian array $\mathscr{E} = \{\epsilon_p(i), \ i \in [p]\}$ with standard Normal marginal. We shall assume that $\mathscr{E}$ is (uniformly) relatively stable, so that in particular,

$$\frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i) =: \frac{M_p}{u_p} \xrightarrow{P} 1,$$

as $p \to \infty$, where $u_p := \Phi^{-1}(1 - 1/p)$ is the $(1/p)$-th tail quantile of the standard Normal distribution.

We consider the iid case first and, for clarity, let M_p^* denote the maximum of p independent standard Normal random variables. Suppose that for some $a_p > 0$ and $a_p, \ b_p \in \mathbb{R}$, we have

$$\Phi(a_p^{-1}x + b_p)^p \to \Lambda(x) := \exp\{-e^{-x}\}, \quad \text{as } p \to \infty,$$

for all $x \in \mathbb{R}$. That is, we have

$$a_p(M_p^* - b_p) \xrightarrow{d} \zeta, \quad \text{as } p \to \infty, \tag{3.1}$$

where ζ has the standard Gumbel distribution Λ. The next result is well-known. We give it here since it summarizes and clarifies the possible choices of the normalizing constants a_p and b_p for Eq. 3.1 to hold.

Lemma 3.1 (i) *We have that*

$$\tilde{u}_p(M_p^* - \tilde{u}_p) \xrightarrow{d} \zeta \quad \text{if and only if} \quad p\overline{\Phi}(\tilde{u}_p) \to 1, \tag{3.2}$$

as $p \to \infty$. In this case, $\tilde{u}_p \sim \sqrt{2\log(p)}$ and more precisely

$$\sqrt{2\log(p)}(\tilde{u}_p - u_p^*) \to 0, \quad \text{as } p \to \infty, \tag{3.3}$$

where

$$u_p^* := \sqrt{2\log(p)}\left(1 - \frac{\log(\log(p)) + \log(4\pi)}{4\log(p)}\right). \tag{3.4}$$

(ii) *Relation (3.1) holds if and only if*

$$a_p \sim \sqrt{2\log(p)} \quad \text{and} \quad p\overline{\Phi}(b_p) \to 1.$$

In particular, by part (i), we have that Eq. 3.1 *holds with $a_p := b_p$ and* Eq. 3.3 *holds with $\tilde{u}_p := b_p$.*

Proof Part (i). Observe that by the Mill's ratio (cf Lemma 5.1), $p\overline{\Phi}(\tilde{u}_p) \to 1$ is equivalently expressed as follows:

$$p\overline{\Phi}(\tilde{u}_p) \sim p\frac{\phi(\tilde{u}_p)}{\tilde{u}_p} \to 1, \quad \text{as } p \to \infty,$$

where $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$ is the standard Normal density. By taking logarithms, the above asymptotic relation is equivalent to having

$$\log(p) - \frac{\widetilde{u}_p^2}{2} - \log(\widetilde{u}_p) - \frac{1}{2}\log(2\pi) \to 0. \tag{3.5}$$

We first prove the 'if' direction of part (i). Suppose that $p\overline{\Phi}(\widetilde{u}_p) \to 1$, or equivalently, Eq. 3.5 holds. Then, one necessarily has $\widetilde{u}_p \to \infty$. It is easy to see that Eq. 3.1 holds with $a_p := \widetilde{u}_p$ and $b_p := \widetilde{u}_p$, provided that, for all $x \in \mathbb{R}$,

$$\Phi\left(\widetilde{u}_p + \frac{x}{\widetilde{u}_p}\right)^p \to \Lambda(x), \quad \text{as } p \to \infty. \tag{3.6}$$

The latter, upon taking logarithms and using the fact that $\log(1+z) \simeq z$, as $z \to 0$, is equivalent to having

$$p\overline{\Phi}\left(\widetilde{u}_p + \frac{x}{\widetilde{u}_p}\right) \to -\log(\Lambda(x)) = e^{-x}. \tag{3.7}$$

To prove that Eq. 3.7 holds, as argued above, using the Mill's ratio, it is equivalent to verify that

$$A_p := \log(p) - \frac{1}{2}\left(\widetilde{u}_p + x/\widetilde{u}_p\right)^2 - \log\left(\widetilde{u}_p + x/\widetilde{u}_p\right) - \frac{1}{2}\log(2\pi) \to -x,$$

as $p \to \infty$. Note that, upon expanding the square and manipulating the logarithm, we obtain

$$A_p = \log(p) - \frac{\widetilde{u}_p^2}{2} - \log(\widetilde{u}_p) - \frac{1}{2}\log(2\pi) - x - x^2/(2\widetilde{u}_p^2) - \log(1 + x/\widetilde{u}_p^2).$$

In view of Eq. 3.5 and the fact that $\widetilde{u}_p \to \infty$, we obtain that $A_p \to -x$, which yields Eq. 3.7 and completes the proof of the 'if' direction of part (i).

Now, to show the 'only if' direction of part (i), suppose that Eq. 3.1 holds with $a_p = b_p := \widetilde{u}_p$, or, equivalently (3.6) holds. By letting $x = 0$ in Relation (3.6), we see that $\widetilde{u}_p \to \infty$, and then, upon taking logarithms, necessarily $p\overline{\Phi}(\widetilde{u}_p) \to 1$, which completes the proof of Eq. 3.2.

We now show Eq. 3.3. First, one can directly verify that Eq. 3.5 holds with $\widetilde{u}_p$ replaced by u_p^* in Eq. 3.4. This, as argued above, is equivalent to $p\overline{\Phi}(u_p^*) \to 1$. Suppose now that, for another sequence $\widetilde{u}_p$, we have $p\overline{\Phi}(\widetilde{u}_p) \to 1$. Then, by the shown equivalence in Eq. 3.2,

$$u_p^*(M_n^* - u_p^*) \xrightarrow{d} \zeta \quad \text{and} \quad \widetilde{u}_p(M_n^* - \widetilde{u}_p) \xrightarrow{d} \zeta.$$

Thus, the convergence of types theorem (see, *e.g.*,Theorem 14.2 in Billingsley 1995) yields

$$u_p^* \sim \widetilde{u}_p \quad \text{and} \quad u_p^*(u_p^* - \widetilde{u}_p) \to 0.$$

The last convergence implies the claim of part (ii) since in view of Eq. 3.4, we have $u_p^* \sim \sqrt{2\log(p)}$.

Part (ii) is a direct consequence of the convergence to types theorem, as argued in the proof of part (i). □

The following result characterizes the optimal rate of concentration under an additional distributional convergence assumption, which holds under the Berman condition for e.g. the case of stationary time series.

Proposition 3.1 *Suppose that $\mathscr{E}$ is a dependent triangular Gaussian array, such that*

$$\zeta_p := a_p(M_p - b_p) \xrightarrow{d} \zeta, \quad as\ p \to \infty, \tag{3.8}$$

for some non-degenerate random variable ζ, with the same constants as in the iid case (3.1). Suppose also that $\mathbb{P}(\zeta < x) > 0$ and $\mathbb{P}(\zeta > x) > 0$ for all $x \in \mathbb{R}$.

Let now the sequence $\delta_p \to 0$, be an upper bound on the rate of concentration, i.e., we have

$$\mathbb{P}\left(\left|\frac{M_p}{a_p} - 1\right| > \delta_p\right) \to 0, \quad p \to \infty. \tag{3.9}$$

The following two statements hold.

(a) *When $\limsup_{p\to\infty} a_p|b_p - a_p| < \infty$, Relation (3.9) holds if and only if*

$$\delta_p \gg \frac{1}{a_p^2} + \left|\frac{b_p}{a_p} - 1\right| =: \delta_p^{opt}. \tag{3.10}$$

(b) *When $\limsup_{p\to\infty} a_p|b_p - a_p| = \infty$, Relation (3.9) holds if and only if*

$$\liminf_{p\to\infty}\left[\frac{\delta_p}{\delta_p^{opt}} - 1\right](1 + a_p|b_p - a_p|) = \infty. \tag{3.11}$$

Proof (a) We will start with the "if" direction. Relation (3.8) implies that

$$\frac{1}{a_p}M_p = \frac{\zeta_p}{a_p^2} + \frac{b_p}{a_p}.$$

Since by assumption the constants a_p and b_p are the same as in the iid case (3.1), Lemma 3.1 entails that $b_p \sim a_p \sim \sqrt{2\log(p)}$. Hence

$$\frac{1}{a_p}M_p - 1 = \frac{\zeta_p}{a_p^2} + \left(\frac{b_p}{a_p} - 1\right) \to 0, \tag{3.12}$$

which shows that the distributional limit in Eq. 3.8 entails concentration of the maxima M_p/a_p to 1. Relations (3.10) and (3.12), however imply that

$$\left|\frac{M_p}{a_p} - 1\right| = o_P(\delta_p),$$

which entails (3.9) by Slutsky (or also Lemma 3.2, below.)

Now, for the converse direction, suppose that Eq. 3.9 holds for some $\delta_p \gg \delta_p^{opt}$. This means that we can find a subsequence $p(n)$ so that $\delta_{p(n)} \leq c \cdot \delta_{p(n)}^{opt}$, $\forall n \in \mathbb{N}$, for a positive constant c that does not depend on n. In view of Eq. 3.9, this would mean that

$$\theta_n := \mathbb{P}\left(\left|\frac{M_{p(n)}}{a_{p(n)}} - 1\right| > c\delta_{p(n)}^{opt}\right) \to 0, \quad n \to \infty.$$

Moreover, since $\limsup_{p\to\infty} a_p |b_p - a_p| < \infty$, and $a_p > 0$, the sequence $(a_p |b_p - a_p|)_{p=1}^{\infty}$ is bounded. Namely, there exists $M > 0$, such that $0 \le a_p |b_p - a_p| \le M$, for all $p \in \mathbb{N}$. However, we have that

$$
\begin{aligned}
\theta_n &\ge \mathbb{P}\left(\frac{M_{p(n)}}{a_{p(n)}} - 1 > c\delta_{p(n)}^{opt}\right) = \mathbb{P}\left(\frac{\zeta_{p(n)}}{a_{p(n)}^2} + \frac{b_{p(n)}}{a_{p(n)}} - 1 > \frac{c}{a_{p(n)}^2} + c\left|\frac{b_{p(n)}}{a_{p(n)}} - 1\right|\right) \\
&= \mathbb{P}\left(\zeta_{p(n)} + a_{p(n)}(b_{p(n)} - a_{p(n)}) - c|a_{p(n)}(b_{p(n)} - a_{p(n)})| > c\right) \\
&\ge \mathbb{P}\left(\zeta_{p(n)} - (c+1)a_{p(n)}|b_{p(n)} - a_{p(n)}| > c\right) \\
&\ge \mathbb{P}\left(\zeta_{p(n)} > c + (c+1)a_{p(n)}|b_{p(n)} - a_{p(n)}|\right) \\
&\ge \mathbb{P}(\zeta_{p(n)} > c + (c+1)M) \\
&\to \mathbb{P}(\zeta > c + (c+1)M) > 0,
\end{aligned}
$$

where the last convergence holds because $\zeta_{p(n)} \xrightarrow{d} \zeta$. This is a contradiction and the proof is complete.

(b) We have that

$$
\begin{aligned}
\mathbb{P}\left(\left|\frac{M_p}{a_p} - 1\right| > \delta_p\right) &= \mathbb{P}\left(a_p |M_p - a_p| > \delta_p a_p^2\right) \\
&= \mathbb{P}\left(|\zeta_p + a_p(b_p - a_p)| > \delta_p a_p^2\right) \\
&= \mathbb{P}\left(\zeta_p < -\delta_p a_p^2 - a_p(b_p - a_p)\right) \\
&\quad + \mathbb{P}\left(\zeta_p > \delta_p a_p^2 - a_p(b_p - a_p)\right) \\
&=: A(p) + B(p).
\end{aligned}
$$

Note, however, that Eq. 3.9 entails that both $A(p)$ and $B(p)$ vanish to 0, as $p \to \infty$. This in turn means that

$$
\liminf_{p\to\infty}(\delta_p a_p^2 - a_p(b_p - a_p)) = \infty \quad \text{and} \quad \liminf_{p\to\infty}(\delta_p a_p^2 + a_p(b_p - a_p)) = \infty,
$$
(3.13)

because of the distributional convergence (3.8). We will work with $B(p)$. The result for $A(p)$ can be obtained by similar arguments. At first, for $B(p)$ to vanish to 0, we do need $\delta_p a_p^2 > a_p(b_p - a_p)$ eventually. Suppose that $\liminf_{p\to\infty}(\delta_p a_p^2 - a_p(b_p - a_p)) = c < \infty$, where $c \ge 0$. This would mean that there is a subsequence $p(n)$ such that

$$
\delta_{p(n)}a_{p(n)}^2 - a_{p(n)}(b_{p(n)} - a_{p(n)}) \to c, \quad p \to \infty.
$$

But then,

$$
B(p(n)) = \mathbb{P}\left(\zeta_{p(n)} > \delta_{p(n)}a_{p(n)}^2 - a_{p(n)}(b_{p(n)} - a_{p(n)})\right) \to \mathbb{P}(\zeta > c) > 0,
$$

which contradicts the fact that $B(p) \to 0$, as $p \to \infty$.

Finally, note that Eq. 3.13 is equivalent to $\liminf_{p\to\infty}(\delta_p a_p^2 - a_p|b_p - a_p|) = \infty$, which with straightforward algebra can be expressed as Eq. 3.11. Indeed,

$$\delta_p a_p^2 - a_p|b_p - a_p| = a_p^2 \left[\delta_p - \left|\frac{b_p}{a_p} - 1\right|\right] = a_p^2 \left[\delta_p - \delta_p^{opt}\right] + 1$$

$$= a_p^2 \delta_p^{opt}\left[\frac{\delta_p}{\delta_p^{opt}} - 1\right] + 1$$

$$= \left[\frac{\delta_p}{\delta_p^{opt}} - 1\right](1 + a_p|b_p - a_p|) + 1,$$

which completes the proof.

$\square$

Remark 3.1 (On the optimality of the rate δ_p^{opt}) The rate δ_p^{opt} can be viewed as "the" optimal rate of concentration in Eq. 3.9 in the sense of Eqs. 3.10 and 3.11. As pointed out by an anonymous referee, the distributional convergence in Eq. 3.8 (whenever it takes place) is much more informative than a simple concentration of maxima type convergence. Specifically, by Lemma 3.1 (ii), one can take $u_p = a_p = b_p$, and in this case Relation (3.12) implies that $1/a_p^2 \propto 1/\log(p)$ is both an upper and lower bound on the rate of concentration. That is, the rate $\delta_p^{opt} = 1/a_p^2 \propto 1/\log(p)$ cannot be improved and in this sense is the optimal rate at which the maxima can concentrate. The rate of concentration, though, does depend on the choice of the normalization sequence u_p. We elaborate on this point next.

The role of the sequence u_p It is well-known that under quite substantial dependence, the convergence in distribution (3.8) holds, with the *same constants* as in the independent case. For example, suppose that $\epsilon_p(i) = Z(i)$, $i \in \mathbb{Z}$ come from a stationary Gaussian time series, which satisfies the so-called *Berman condition* (Berman 1964):

$$\mathrm{Cov}(Z(k), Z(0)) = o\left(\frac{1}{\log(k)}\right), \quad \text{as } k \to \infty.$$

Notice, by Lemma 3.1 (ii), however, we also have $\widetilde{\zeta}_p := b_p(M_p - b_p) \xrightarrow{d} \zeta$, and

$$\frac{1}{b_p}M_p - 1 = \frac{\widetilde{\zeta}_p}{b_p^2} = \mathcal{O}_P\left(\frac{1}{\log(p)}\right). \tag{3.14}$$

Compare Relations (3.12) and (3.14). Since $a_p \sim b_p \sim \sqrt{2\log(p)}$, from Eq. 3.14, we have that the rate of concentration of M_p relative to the sequence b_p is $1/\log(p)$. On the other hand, while the first term in the right-hand side of Eq. 3.12 is of order $1/\log(p)$ the presence of the second term can only make the rate of concentration therein *slower*. Indeed, this is formally established in Lemma 3.2. To gain some more

intuition that the poor choice of a sequence a_p can lead to a slower rate of concentration, suppose that $a_p = b_p/(1 + g(p))$, for an arbitrary sequence $g(p) > -1$, such that $g(p) \to 0$. Then, by Eq. 3.12,

$$\frac{1}{a_p} M_p - 1 = \frac{\zeta_p}{a_p^2} + g(p).$$

One can take $g(p) \to 0$ arbitrarily slow. Finally, as a more concrete example, one typically uses $b_p := u_p^* = \sqrt{2\log(p)}(1 - (\log(\log(p)) + \log(4\pi))/4\log(p))$ and $a_p := \sqrt{2\log(p)}$. It is easily seen that $b_p = a_p(1 + g(p))$, where

$$g(p) = -\frac{\log(\log(p)) + \log(4\pi)}{4\log(p)} \propto \frac{\log(\log(p))}{\log(p)}.$$

This shows that, in particular, in the case of iid maxima (as well as in the general case where Eq. 3.8 holds) the normalization $\sqrt{2\log(p)}$ *does not lead* to the optimal rate, since

$$\frac{1}{\sqrt{2\log(p)}} M_p^* - 1 \propto_P \frac{\log(\log(p))}{\log(p)},$$

where $\xi_p \propto_P \eta_p$ means that $\xi_p/\eta_p \to c$ in probability, for some positive constant c.

The optimal rate is $1/\log(p)$ and it is obtained by normalizing with any sequence b_p such that $p\overline{\Phi}(b_p) \to 1$. This follows from the next simple result, which shows that the rate of concentration in Eq. 3.12 is the slower of the rates $1/a_p^2$ and $(b_p - a_p)/a_p$.

Lemma 3.2 *Suppose that for some random variables ζ_p, we have $\zeta_p \xrightarrow{d} \zeta$, as $p \to \infty$, where ζ is a non-constant random variable. Then, for all sequences α_p and β_p, we have*

$$\alpha_p \zeta_p + \beta_p \xrightarrow{P} 0 \iff |\alpha_p| + |\beta_p| \longrightarrow 0.$$

That is, the rate of $\alpha_p \zeta_p + \beta_p$ is always the slower of the rates of $\{\alpha_p\}$ and $\{\beta_p\}$.

Proof The '$\Leftarrow$' direction follows from Slutsky. To prove '$\Rightarrow$', it is enough to show that for every $p(n) \to \infty$, there is a further sub-sequence $q(n) \to \infty$, $\{q(n)\} \subset \{p(n)\}$, such that

$$|\alpha_{q(n)}| + |\beta_{q(n)}| \longrightarrow 0.$$

In view of Skorokhod's representation theorem (Theorem 6.7, page 70 in Billingsley 2013), we may suppose that $\zeta_p^* \to \zeta^*$, with probability one, where $\zeta_p^* \overset{d}{=} \zeta_p$ and $\zeta^* \overset{d}{=} \zeta$. Also, assuming that $\alpha_{p(n)} \zeta_{p(n)}^* + \beta_{p(n)} \to 0$, in probability, implies that there is a further sub-sequence $q(n) \to \infty$, such that

$$\alpha_{q(n)} \zeta_{q(n)}^*(\omega) + \beta_{q(n)} \to 0, \quad \text{as } q(n) \to \infty, \tag{3.15}$$

for P-almost all ω. Since also $\zeta_{q(n)}^*(\omega) \to \zeta^*(\omega)$, for P-almost all ω, and since ζ^* is non-constant, we have $\zeta_{q(n)}^*(\omega_i) \to \zeta^*(\omega_i)$, $i = 1, 2$ for some $\zeta^*(\omega_1) \neq \zeta^*(\omega_2)$.

Thus, by subtracting two instances of Relation (3.15) corresponding to $\omega = \omega_1$ and $\omega = \omega_2$, we obtain

$$\alpha_{q(n)}(\zeta_{q(n)}^*(\omega_1) - \zeta_{q(n)}^*(\omega_2)) \to 0,$$

which since $(\zeta^*_{q(n)}(\omega_1) - \zeta^*_{q(n)}(\omega_2)) \to \zeta^*(\omega_1) - \zeta^*(\omega_2) \neq 0$, implies $\alpha_{q(n)} \to 0$. This, in view of Eq. 3.15 yields $\beta_{q(n)} \to 0$, and completes the proof. $\qquad\square$

Remark 3.2 The above considerations establish the optimal rate of concentration of the maxima $M_p = \max_{i \in [p]} \epsilon_p(i)$, whenever the limit in distribution (3.8) holds. We have shown that this optimal rate is $1/\log(p)$ and is in fact obtained, when considering M_p/u_p, for $p\overline{\Phi}(u_p) \sim 1$. The rate of concentration of $M_p/\sqrt{2\log(p)}$ is $\log(\log(p))/\log(p)$, which is only slightly sub-optimal.

On the other hand, as we know by Theorem 2.1, uniform relative stability is equivalent to UDD and hence the concentration of maxima phenomenon takes place even if Eq. 3.8 fails to hold. At this point, we do not know what is the optimal rate in general. In Section 4, we provide upper bounds on this rate. We conjecture, however, the presence of more severe dependence can only lead to slower rates of concentration and in particular the optimal rate of concentration for UDD arrays cannot be faster than $1/\log(p)$—the one for independent maxima.

Conjecture 1 *Let $\mathscr{E}$ be a Gaussian URS array. Relation* (4.3) *implies* $\delta_p \gg 1/\log(p)$.

4 Rates of uniform relative stability

4.1 Gaussian arrays

Throughout Sections 4 and 5, $\mathscr{E} = \{\epsilon_p(i), \ i \in [p]\}$ will be a Gaussian array with standard Normal marginals, unless stated otherwise. We shall also assume that $\mathscr{E}$ is URS. For simplicity of notation and without loss of generality we will work with $S_p = [p]$ (see Remark 4.2). We will obtain upper bounds on the rate, i.e., sufficient conditions on the dependence structure of $\mathscr{E}$, which ensure certain rates. These results are of independent interest and will find concrete applications in Section 4.2, where conditions ensuring the URS of functions of Gaussian arrays are established.

The following definition is an ancillary tool for the comparison of the rates of two vanishing sequences and introduces some notation for this purpose.

Definition 4.1 Let $(\alpha_p)^\infty_{p=1}$ and $(\beta_p)^\infty_{p=1}$ be two positive sequences converging to 0. We will say that α_p is of lower order than β_p (or slower than β_p), denoted by $\alpha_p \gg \beta_p$, if $\beta_p/\alpha_p \to 0$, as $p \to \infty$, i.e., $\beta_p = o(\alpha_p)$.

The next theorem constitutes the main result of this paper.

Theorem 4.1 *Consider a UDD Gaussian triangular array $\mathscr{E} = \{\epsilon_p(i), i \in [p]\}$ with standard Normal marginals and let $N_\mathscr{E}(\tau)$ be as in Definition 2.3. Let $\tau(p) \to 0$ be such that*

$$\alpha(p) := \log N_\mathscr{E}(\tau(p)) / \log(p) \to 0, \quad as \ p \to \infty. \tag{4.1}$$

Then, for all $\delta_p > 0$ such that

$$\delta_p \gg \alpha(p) + \tau(p) + \frac{1}{\log(p)}, \tag{4.2}$$

we have

$$\mathbb{P}\left(\left|\frac{\max_{i \in [p]} \epsilon_p(i)}{u_p} - 1\right| > \delta_p\right) \to 0, \quad as \ p \to \infty. \tag{4.3}$$

Here u_p is defined as in Eq. 2.2 *taking $F = \Phi$, the cumulative distribution function of standard Normal distribution.*

The proof of Theorem 4.1 depends on a number of technical results, which will be presented and proved in Section 5. In order to make the proof easier for the reader to follow, we postpone its demonstration until Section 5. We proceed next with several comments and examples.

Remark 4.1 Note that in Theorem 4.1 the covariance structure of $\mathscr{E}$ appears only through $N_{\mathscr{E}}(\tau)$. The collection $\{N_{\mathscr{E}}(\tau), \ \tau \in (0, 1)\}$ constitutes a collection of uniform upper bounds on the number of covariances in each row of the triangular array $\mathscr{E}$ that exceed the threshold τ. This means that the ordering of the p random variables in each row of $\mathscr{E}$ is irrelevant.

Remark 4.2 The support recovery results of Gao and Stoev (2020) require URS in the sense of Eq. 2.5 for a subsequence $S_p \subset [p]$, with $|S_p| \to \infty$. By the previous remark, upon relabelling the triangular array $\mathscr{E}$, Theorem 4.1 applies in this setting with p replaced by $|S_p|$, and entails rates on the convergence in Eq. 2.5.

The preceding Theorem 4.1 gives us an upper bound on the rate at which the convergence in Eq. 2.5 takes place for a UDD Gaussian array $\mathscr{E}$. Observe that this bound depends crucially on the covariance structure of $\mathscr{E}$ through $N_{\mathscr{E}}(\tau)$. This dependence will be illustrated in the following examples, where the upper bound stated in Eq. 4.2 is obtained for three specific covariance structures.

Example 4.1 The iid case and optimality of the rate bounds.
Suppose that all $\epsilon_p(j)$'s are iid. Then, we can pick $\tau(p) = 0$ or $\tau < 1$ vanishing to 0 arbitrarily fast, and we would have that $N_{\mathscr{E}}(\tau) = 1$, because of the strict inequality in Eq. 2.6. This implies that $\alpha(p) = \log(N_{\mathscr{E}}(\tau))/\log(p) = 0$. Thus, in this case, the upper bound in Eq. 4.2 becomes $1/\log(p)$. Observe that this rate matches the optimal rate in Conjecture 1.

Example 4.2 Power-law covariance decay.
Consider, first, the simple case where $\mathscr{E}$ comes from a stationary Gaussian time series, $\epsilon_p(\kappa) = \epsilon(\kappa)$, with auto-covariance

$$\rho(\kappa) = \text{Cov}(\epsilon(\kappa), \epsilon(0)) \propto \kappa^{-\gamma}, \quad \gamma > 0. \tag{4.4}$$

Then, the classic Berman condition $\rho(\kappa) = o(1/\log(\kappa))$ holds and as shown in the discussion after Proposition 3.1, the optimal rate in Eq. 2.5 is $1/\log(p)$.

In this example, we will demonstrate that our result [Theorem 4.1] leads to the nearly optimal rate $\log(\log(p))/\log(p)$. As in the previous remark, we see that this is in fact the optimal rate if u_p in Eq. 2.5 is replaced by $\sqrt{2\log(p)}$. (See Section 3). Note, however, that our arguments apply in greater generality and do not depend on the stationarity assumption. Indeed, assume that $\mathscr{E}$ is a general Gaussian triangular array such that (UDD$'$) of Gao and Stoev (2020) holds, i.e.,

$$\left|\mathrm{Cov}(\epsilon_p(i), \epsilon_p(j))\right| \le c \left|\pi_p(i) - \pi_p(j)\right|^{-\gamma} \tag{4.5}$$

for suitable permutations π_p of $\{1, \dots, p\}$, where c does not depend on p. (Note that Eq. 4.5 entails Eq. 4.4 for $\pi_p = id$, where id is the identity permutation.) Then, one can readily show that $N_{\mathscr{E}}(\tau) = \mathscr{O}(\tau^{-1/\gamma})$, as $\tau \to 0$. Thus,

$$\alpha(p) = \frac{\log(N_{\mathscr{E}}(\tau))}{\log(p)} \propto \frac{\log(\tau^{-1/\gamma})}{\log(p)} = -\frac{\log(\tau)}{\gamma \log(p)}.$$

Using this $\alpha(p)$, the upper bound on the rate in Theorem 4.1 becomes

$$\alpha(p) + \tau(p) + \frac{\log(\log(p))}{\log(p)} \propto -\frac{\log(\tau)}{\log(p)} + \tau(p) + \frac{1}{\log(p)} \asymp -\frac{\log(\tau)}{\log(p)} + \tau(p). \tag{4.6}$$

This is minimized by taking $\tau(p) = 1/\log(p)$ in Eq. 4.6 and the upper bound on the rate becomes

$$\alpha(p) + \tau(p) + \frac{1}{\log(p)} \propto \frac{\log(\log(p))}{\log(p)} + \frac{1}{\log(p)} \asymp \frac{\log(\log(p))}{\log(p)}.$$

Recall that in the case when $\mathscr{E}$ has iid components, the optimal rate of concentration of the maxima is $1/\log(p)$ and in fact it becomes $\log(\log(p))/\log(p)$ when one uses the normalization $\sqrt{2\log(p)}$ in place of u_p. Therefore, this example shows that under mild power-law type covariance decay conditions, Gaussian triangular arrays continue to concentrate at the nearly optimal rates for the iid setting.

Example 4.3 Logarithmic covariance decay.

Following suit from Example 4.2, we consider first the case where the errors come from a stationary time series with auto-covariance

$$\rho(\kappa) = \mathrm{Cov}(\epsilon(\kappa), \epsilon(0)) \propto (\log(\kappa))^{-\nu}, \quad \text{as } \kappa \to \infty, \tag{4.7}$$

for some $\nu > 0$. Note that for $0 < \nu < 1$, the Berman condition $\rho(\kappa) = o(1/\log(\kappa))$ is no longer satisfied and the results from Section 3 cannot be applied to establish the optimal rate in Eq. 2.5. Using Theorem 4.1, we will see that an upper bound on this rate is $\delta_p^\star := (\log(p))^{-\frac{\nu}{\nu+1}}$.

Indeed, consider the more general case where $\mathscr{E}$ is a Gaussian triangular array, such that (UDD$'$) of Gao and Stoev (2020) holds, i.e.,

$$\left|\mathrm{Cov}(\epsilon_p(i), \epsilon_p(j))\right| \le c \left(\log\left(\left|\pi_p(i) - \pi_p(j)\right|\right)\right)^{-\nu}, \tag{4.8}$$

for suitable permutations π_p of $\{1, \ldots, p\}$ and c does not depend on p. Again, note that Eq. 4.8 implies Eq. 4.7 for the identity permutation. One can show that in this case $N_{\mathscr{E}}(\tau) = \mathscr{O}\left(e^{\tau^{-1/\nu}}\right)$, as $\tau \to 0$ and thus,

$$\alpha(p) = \frac{\log(N_{\mathscr{E}}(\tau))}{\log(p)} \propto \frac{\log\left(e^{\tau^{-1/\nu}}\right)}{\log(p)} = \frac{1}{\tau^{1/\nu} \log(p)}, \quad \text{as } p \to \infty.$$

To find the best bound on the rate in the context of Eq. 4.2 we minimize

$$\alpha(p) + \tau(p) + \frac{1}{\log(p)} \propto \frac{1}{\tau^{1/\nu} \log(p)} + \tau + \frac{1}{\log(p)},$$

with respect to τ. Considering p fixed, basic calculus gives us that the r.h.s. is minimized for $\tau(p) = (\nu \log(p))^{-\frac{\nu}{\nu+1}}$. With this choice of τ the fastest upper bound from Theorem 4.1 becomes

$$\left[\nu^{-\frac{\nu}{\nu+1}} + \nu^{-\frac{1}{\nu+1}}\right] \cdot (\log(p))^{-\frac{\nu}{\nu+1}} + \frac{1}{\log(p)} \propto (\log(p))^{-\frac{\nu}{\nu+1}}.$$

It only remains to show that the choice of τ actually allows us to pick $N_{\mathscr{E}}(\tau) = \mathscr{O}\left(e^{\tau^{-1/\nu}}\right)$. A sufficient condition would be $p \geq \tilde{c} \cdot e^{\tau^{-1/\nu}}$ for a suitably chosen constant $\tilde{c}$ not depending on either p or τ. Substituting $\tau = (\nu \log(p))^{-\frac{\nu}{\nu+1}}$, we equivalently need

$$p \geq \tilde{c} \cdot e^{(\nu \log(p))^{\frac{1}{\nu+1}}}.$$

It is readily checked, by taking logarithms in both sides, that this holds for p sufficiently large and thus, the fastest upper bound for this kind of dependence structure is $(\log(p))^{-\frac{\nu}{\nu+1}}$.

Observe that as $\nu \to \infty$ this upper bound approaches asymptotically the optimal rate $1/\log(p)$ achieved under the Berman condition (see Section 3). Our results yield, however, an upper bound on the rate of concentration in Eq. 2.5 for the case $0 < \nu < 1$, where the Berman condition does not hold.

4.2 Functions of Gaussian arrays

The main motivation behind the work in this section is to determine when the concentration of maxima property is preserved under transformations. Specifically, consider the triangular array

$$\mathscr{H} = \{\eta_p(j) = f(\epsilon_p(j)), \ j \in [p], \ p \in \mathbb{N}\}, \tag{4.9}$$

where $\mathscr{E} = \{\epsilon_p(j), \ j \in [p], \ p \in \mathbb{N}\}$ is a Gaussian triangular array with standard Normal marginals.

Given that Eq. 4.3 holds, our goal is to find bounds on a sequence $d_p \downarrow 0$, such that

$$\mathbb{P}\left(\left|\frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1\right| > d_p\right) \to 0, \quad \text{as } p \to \infty, \tag{4.10}$$

where $v_p = f(u_p)$ and u_p is as in Eq. 2.2. We first address the case of monotone non-decreasing transformations.

Proposition 4.1 *Asssume that f is a non-decreasing differentiable and eventually strictly increasing function, with $\lim_{x \to \infty} f(x) \neq 0$ and the derivative $f'(x)$ is either eventually increasing or eventually decreasing as $x \to \infty$. If Eq. 4.3 holds with some $\delta_p > 0$, then Eq. 4.10 holds provided that*

$$d_p \geq d_p^{\star} := \frac{u_p \delta_p \max \left\{ |f'(u_p(1 - \delta_p))|, |f'(u_p(1 + \delta_p))| \right\}}{|f(u_p)|}. \tag{4.11}$$

Proof Since $u_p \uparrow \infty$, by the monotonicity of f and the fact that it is eventually strictly increasing, one can show that $f(u_p) = v_p = F_\eta^{\leftarrow}(1 - 1/p)$, for p large enough. We start by noticing that

$$\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| = \left| \frac{\max_{j \in [p]} f(\epsilon_p(j)) - f(u_p)}{f(u_p)} \right|$$

$$= \left| \frac{f\left(\max_{j \in [p]} \epsilon_p(j) \right) - f(u_p)}{f(u_p)} \right|, \tag{4.12}$$

where the second equality follows by the monotonicity of f.

Now recall that f is differentiable. By the Mean Value Theorem, there exists a possibly random θ_p between u_p and $\max_{j \in [p]} \epsilon_p(j)$, such that

$$\left| \frac{f\left(\max_{j \in [p]} \epsilon_p(j) \right) - f(u_p)}{f(u_p)} \right| = \left| \frac{1}{f(u_p)} f'(\theta_p) \left(\max_{j \in [p]} \epsilon_p(j) - u_p \right) \right|. \tag{4.13}$$

Combining Eqs. 4.12 and 4.13, we obtain

$$\mathbb{P}\left(\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| > d_p \right) = \mathbb{P}\left(\left| \frac{u_p f'(\theta_p)}{f(u_p)} \right| \cdot \left| \frac{\max_{j \in [p]} \epsilon_p(j)}{u_p} - 1 \right| > d_p \right)$$

$$= \mathbb{P}\left(\left| \frac{\max_{j \in [p]} \epsilon_p(j)}{u_p} - 1 \right| > \frac{d_p |f(u_p)|}{u_p |f'(\theta_p)|} \right),$$

where the second equality follows from the fact that $f'(\theta_p) \neq 0$ over the event of interest, since $d_p > 0$. This shows that for any non-negative sequence δ_p vanishing to 0, such that Eq. 4.3 holds, we have that

$$\mathbb{P}\left(\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| > \tilde{d}_p \right) \to 0, \quad \text{as } p \to \infty, \tag{4.14}$$

where

$$\tilde{d}_p := \frac{u_p \delta_p |f'(\theta_p)|}{|f(u_p)|}. \tag{4.15}$$

Now, we know by Eq. 4.3 that

$$\left| \theta_p - u_p \right| \leq \left| \max_{j \in [p]} \epsilon_p(j) - u_p \right| \leq u_p \delta_p$$

with probability going to 1, as $p \to \infty$. This implies that

$$\mathbb{P}\left(u_p(1 - \delta_p) \leq \theta_p \leq u_p(1 + \delta_p) \right) \to 1, \quad \text{as } p \to \infty,$$

In turn, by the eventual monotonicity of f', the last convergence implies that

$$\mathbb{P}\left(\left|f'(\theta_p)\right| \leq \max\left\{\left|f'(u_p(1-\delta_p))\right|, \left|f'(u_p(1+\delta_p))\right|\right\}\right) \to 1, \quad \text{as } p \to \infty,$$

and equivalently

$$\mathbb{P}\left(\tilde{d}_p \leq d_p^\star\right) \to 1, \quad \text{as } p \to \infty. \tag{4.16}$$

By Eqs. 4.15 and 4.16 we conclude that Eq. 4.14 holds with $\tilde{d}_p$ substituted by $d_p^\star$. This shows that $d_p^\star$ is an upper bound of the optimal rate of concentration, i.e., Eq. 4.11 implies Eq. 4.10. □

A typical and very important case where Proposition 4.1 applies is when the array $\mathscr{E}$ undergoes an exponential transformation, illustrated in the following example.

Example 4.4 Let $\mathscr{E}$ be as in Proposition 4.1 and consider

$$\mathscr{H}_\mathscr{E} = \left\{\eta_p(j) := e^{\epsilon_p(j)}, \ j \in [p], \ p \in \mathbb{N}\right\}, \tag{4.17}$$

which is a triangular array with lognormal marginal distributions. This is sometimes referred to as the multivariate lognormal model (Halliwell 2015). Let δ_p be such that Eq. 4.3 holds. Then, an immediate application of Proposition 4.1 shows that as long as $u_p\delta_p \to 0$, an upper bound on the rate of convergence in Eq. 4.10 is

$$d_p^\star = u_p\delta_p e^{u_p\delta_p} \sim u_p\delta_p \sim \delta_p\sqrt{2\log(p)}.$$

That is, lognormal arrays can have relatively stable maxima, provided that the underlying maxima of the Gaussian array concentrate at a rate $\delta_p = o\left(1/\sqrt{\log(p)}\right)$.

Popular models like the ones with χ_1^2 marginals can be obtained from Proposition 4.1 with the monotone transformation $f(x) := F^{-1}(\Phi(x))$, where F is the cdf of the desired distribution. The classic multivariate χ_1^2- models, however, are obtained by squaring the elements of the Gaussian array, i.e., via the non-monotone transformation $f(x) = x^2$. Such models are addressed in the next result.

Corollary 4.1 *Let all the assumptions of Proposition 4.1 hold and let $d_p^\star$ be defined as before. Assume now that f is an even ($f(x) = f(-x)$) differentiable and eventually strictly increasing function, with $\lim_{x\to\infty} f(x) \neq 0$. Assume also that f is monotone non-decreasing on $(0, \infty)$. Then, the conclusion Eq. 4.11 still holds.*

Proof We start by observing that

$$\mathbb{P}\left(\left|\frac{\max_{j\in[p]}\eta_p(j)}{f(u_p)} - 1\right| > d_p\right) = \mathbb{P}\left(\left|\frac{\max_{j\in[p]}f(\epsilon_p(j)) - f(u_p)}{f(u_p)}\right| > d_p\right)$$

$$\leq \mathbb{P}\left(\left|\frac{f(\min_{j\in[p]}\epsilon_p(j)) - f(u_p)}{f(u_p)}\right| > d_p\right) + \mathbb{P}\left(\left|\frac{f(\max_{j\in[p]}\epsilon_p(j)) - f(u_p)}{f(u_p)}\right| > d_p\right),$$

$$\tag{4.18}$$

because the symmetry and monotonicity of f on $(0, \infty)$ imply that $\max_{j \in [p]} f(\epsilon_p(j))$ equals either $f\left(\max_{j \in [p]} \epsilon_p(j)\right)$ or $f\left(\min_{j \in [p]} \epsilon_p(j)\right)$.

By Proposition 4.1 we can readily obtain that for $d_p \geq d_p^\star$ the second term of Eq. 4.18 converges to 0. Now, we handle the first term of Eq. 4.18. By the symmetry of f we have that

$$f(\min_{j \in [p]} \epsilon_p(j)) = f(-\min_{j \in [p]} \epsilon_p(j)) = f(\max_{j \in [p]}(-\epsilon_p(j)).$$

Notice that by verifying the equality of the covariance structures, we have

$$\{-\epsilon_p(j), \ j \in [p]\} \overset{d}{=} \{\epsilon_p(j), \ j \in [p]\}.$$

Hence $\max_{j \in [p]}(-\epsilon_p(j)) \overset{d}{=} \max_{j \in [p]} \epsilon_p(j)$, and again by Proposition 4.1 we get that for $d_p \geq d_p^\star$ the first term of Eq. 4.18 also converges to 0. This completes the proof. □

Using Corollary 4.1 we can now treat the multivariate χ^2 model introduced in Dasgupta and Spurrier (1997).

Example 4.5 Let $\mathscr{E}$ be as in Proposition 4.1 and consider

$$\mathscr{H}_{\mathscr{E}} = \left\{\eta_p(j) := \epsilon_p^2(j), \ j \in [p], \ p \in \mathbb{N}\right\},$$

a triangular array with χ_1^2 marginal distributions. Let δ_p be as in Eq. 4.3. Then, a simple application of Corollary 4.1 implies Eq. 4.10, provided

$$d_p \geq d_p^\star = 2\delta_p(1 + \delta_p) \sim 2\delta_p.$$

In contrast to Example 4.4, taking squares does not lead to a slower rate of convergence. Indeed, in Example 4.4 our estimate of the rate is slowed down by a factor of $\sqrt{\log(p)}$, while in the χ^2 case it remains δ_p.

We shall now see that the rate of convergence is not slowed down by any power transformation $x \mapsto x^\lambda$, for any $\lambda > 0$.

Example 4.6 Power-Law Transformations.

Let once again $\mathscr{E}$ be as in Proposition 4.1 and consider the power transformations $f(x) = x^\lambda$, $\lambda > 0$. In the cases where $\lambda \notin \mathbb{N}$, we use the functions $f_1^\lambda(x) = |x|^\lambda$ or $f_2^\lambda(x) = x^{<\lambda>} = \text{sign}(x) \cdot |x|^\lambda$. Note that differentiability at 0 is not needed in any of the proofs, so using f_1^λ does not violate any of the assumptions. Let also δ_p be as in Eq. 4.3, i.e., a rate sequence for the convergence in Eq. 2.5. Then, a suitable application of Proposition 4.1 or Corollary 4.1, shows that an upper bound on the rate of convergence in Eq. 4.10 is

$$d_p^\star = \lambda\delta_p(1 + \delta_p)^{\lambda-1} \sim \lambda\delta_p \qquad \text{or} \qquad d_p^\star = \lambda\delta_p(1 - \delta_p)^{\lambda-1} \sim \lambda\delta_p.$$

In view of Examples 4.1, 4.2 and 4.3, we now show how the rate $d_p^\star \sim \lambda\delta_p$ is affected under different correlation structures of the underlying Gaussian array $\mathscr{E}$.

Recall that in the iid case of Example 4.1 we have that the optimal rate is $\delta_p \gg \delta_p^{\text{opt}} = 1/\log(p)$. This implies that an upper bound on the rate of concentration is

$$d_p^\star \sim \lambda \delta_p \gg \frac{\lambda}{\log(p)}.$$

Moreover, for the power-law covariance decay covariance structure (Example 4.2), we observe that compared to the iid case, the rate of concentration δ_p is scaled by a factor of $\log(\log(p))$. Namely, for the power-law transformations we get that the upper bound is

$$d_p^\star \sim \lambda \delta_p \sim \frac{\lambda \log(\log(p))}{\log(p)}.$$

Finally, we examine the logarithmic covariance decay (Example 4.3). Remember that in this case the rate we have for $\mathscr{E}$ is $\delta_p = (\log(p))^{-\frac{\nu}{\nu+1}}$. This implies that the upper bound of the rate of concentration for the power-law transformations is

$$d_p^\star \sim \lambda \delta_p \sim \frac{\lambda}{(\log(p))^{\frac{\nu}{\nu+1}}}.$$

Observe that in this case, $d_p^\star$ is a valid upper bound aside from the value of ν. We will see in the following Example 4.7, that the same is not true for the exponential power-law transformations.

In the last example of this section, we explore exponential power transformations and how they affect our bounds on the rate of convergence.

Example 4.7 Exponential Power-Law Transformations.

Let $\mathscr{E}$ be as in Proposition 4.1 and consider the exponential power transformations $f(x) = e^{x^\lambda}$, $\lambda > 0$, $\lambda \neq 1$. (Note that $\lambda = 1$ is the lognormal case which we have alredy seen in Example 4.4). In the cases where $\lambda \notin \mathbb{N}$, we use the functions $f_1^\lambda(x) = e^{|x|^\lambda}$ or $f_2^\lambda(x) = e^{x^{<\lambda>}} = e^{\text{sign}(x) \cdot |x|^\lambda}$. Similarly to Example 4.6, differentiability at 0 is not needed in any of the proofs, so using f_1^λ does not violate any of the assumptions. Let also δ_p be as in Eq. 4.3. Then, suitable applications of Proposition 4.1 or Corollary 4.1 show that as long as $u_p^\lambda \delta_p \to 0$, an upper bound on the rate of convergence in Eq. 4.10 is

$$d_p^\star = \lambda u_p^\lambda \delta_p (1 + \delta_p)^{\lambda-1} e^{u_p^\lambda [(1+\delta_p)^\lambda - 1]}, \quad \text{if } \lambda \geq 1$$

and

$$d_p^\star = \lambda u_p^\lambda \delta_p (1 - \delta_p)^{\lambda-1} e^{u_p^\lambda [(1-\delta_p)^\lambda - 1]}, \quad \text{if } 0 < \lambda < 1.$$

In both cases we have $d_p^\star \sim \lambda \delta_p (2\log(p))^{\lambda/2}$, as $p \to \infty$. As a generalization of the lognormal case ($\lambda = 1$), we see that the iid rate δ_p is scaled by a factor of $\left(\sqrt{\log(p)}\right)^\lambda$. This means that this kind of arrays would still have relatively stable maxima, provided that the underlying maxima of the Gaussian array concentrate at a rate $\delta_p = o\left(1/(\log(p))^{\lambda/2}\right)$.

At this point, we examine how the rate $d_p^\star \sim \lambda \delta_p (2\log(p))^{\lambda/2}$ adjusts under the varying covariance structures of $\mathscr{E}$ in Examples 4.1, 4.2 and 4.3. In an analogous

manner to Example 4.6, we get that for the iid case, an upper bound on the rate of concentration is

$$d_p^\star \sim \lambda \delta_p u_p^\lambda \gg 2^{\frac{\lambda}{2}} \lambda \, (\log(p))^{\frac{\lambda}{2}-1},$$

while for the power-law covariance decay covariance structure we obtain

$$d_p^\star \sim \lambda \delta_p u_p^\lambda \sim 2^{\frac{\lambda}{2}} \lambda \, (\log(p))^{\frac{\lambda}{2}-1} \log(\log(p)).$$

In the previous two instances we notice that the covariance structure does not impose any restrictions on the values of λ, in order to guarantee concentration of maxima for the transformed triangular array. This is not the case for the logarithmic covariance decay, since the upper bound becomes

$$d_p^\star \sim \lambda \delta_p u_p^\lambda \sim 2^{\frac{\lambda}{2}} \lambda \, (\log(p))^{\frac{\lambda}{2}-\frac{\nu}{\nu+1}}.$$

The aforementioned $d_p^\star$ is a a sensible upper bound for the rate of concentration in this case, only if $d_p^\star \to 0$, as $p \to \infty$. This is so, when $\nu > \frac{\lambda}{2+\lambda}$. Thus, our results imply that in the lognormal case ($\lambda = 1$), $\nu > \frac{1}{3}$ guarantees that the transformed array is relatively stable.

Remark 4.3 In Conjecture 1, we posit that the fastest rate of convergence for a UDD Gaussian array is bounded above by $1/\log(p)$. Nevertheless, from Example 4.1 for the iid case, our bound in Eq. 4.2 is again $1/\log(p)$. Since $u_p \sim \sqrt{2\log(p)}$, we see that we can get an upper bound on the rate of $f(x) = e^{x^\lambda}$ only for $0 < \lambda < 2$. The range $\lambda \in (0, 2)$ is also natural, because one can show that the transformation $f(x) = e^{x^\lambda}$, for $\lambda \geq 2$, leads to heavy power-law distributed variables $\eta_p(j)$. Heavy-tailed random variables no longer have relatively stable maxima, which makes the question about the rate of concentration of maxima meaningless.

We will end this section with a corollary, readily obtained by the discussion in the end of Example 4.7.

Corollary 4.2 *Suppose that $\mathscr{H} := \{\eta_p(j), \; j \in [p], \; p \in \mathbb{N}\}$ is a multivariate lognormal array as in* Eq. 4.17. *Suppose that*

$$\left| Cov\left(\eta_p(j), \eta_p(k)\right) \right| \leq c \cdot \frac{1}{\left(\log(|\pi_p(j) - \pi_p(k)|)\right)^\nu}, \tag{4.19}$$

for some $\nu > 1/3$, permutations π_p of $\{1, \ldots, p\}$ and a constant c independent of p. Then the array $\mathscr{H}$ is URS.

Proof Let $\mathscr{E} = \{\epsilon_p(j), \; j \in [p], \; p \in \mathbb{N}\}$ be the underlying Gaussian array. Then, we have that $\eta_p(j) = e^{\epsilon_p(j)}$ for every $j \in [p]$. Thus,

$$
\begin{aligned}
\mathrm{Cov}(\eta_p(j), \eta_p(k)) &= \mathrm{Cov}\left(e^{\epsilon_p(j)}, e^{\epsilon_p(k)}\right) \\
&= \mathbb{E}\left(e^{\epsilon_p(j)+\epsilon_p(k)}\right) - \mathbb{E}\left(e^{\epsilon_p(j)}\right) \mathbb{E}\left(e^{\epsilon_p(k)}\right).
\end{aligned} \tag{4.20}
$$

Recall that the moment generating function for a Normal random variable $X \sim N(\mu, \sigma^2)$ is $M(t) = \mathbb{E}\left(e^{tX}\right) = e^{\mu t + \sigma^2 t^2 / 2}$. Since $\epsilon_p(i)$ follow the standard Normal distribution, we have $\epsilon_p(j) + \epsilon_p(k) \sim N(0, 2 + 2\mathrm{Cov}(\epsilon_p(j), \epsilon_p(k)))$, and hence Eq. 4.20 becomes

$$\mathrm{Cov}(\eta_p(j), \eta_p(k)) = e \cdot \left(e^{\mathrm{Cov}(\epsilon_p(j), \epsilon_p(k))} - 1\right). \tag{4.21}$$

In turn, Eq. 4.21 along with Eq. 4.19 implies that

$$\left| e \cdot \left(e^{\mathrm{Cov}(\epsilon_p(j), \epsilon_p(k))} - 1\right) \right| \leq \frac{c}{e} \cdot \frac{1}{\left(\log(|\pi_p(j) - \pi_p(k)|)\right)^v}. \tag{4.22}$$

Using the inequality $|x| \leq e|e^x - 1|$, $x \in [-1, 1]$ in Eq. 4.22, since $\left|\mathrm{Cov}(\epsilon_p(j), \epsilon_p(k))\right| \leq 1$, we finally obtain that

$$\left|\mathrm{Cov}(\epsilon_p(j), \epsilon_p(k))\right| \leq \frac{c}{e} \cdot \frac{1}{\left(\log(|\pi_p(j) - \pi_p(k)|)\right)^v}.$$

The last relation implies that $\mathscr{E}$ has a logarithmic covariance decay covariance structure (see Example 4.3). Combined with the discussion in the end of Example 4.7, the proof is complete. $\qquad\square$

5 Technical proofs

In this section we present the proof of the capstone Theorem 4.1. Recall that we desire to find an upper bound on the rate of positive vanishing sequences δ_p, such that

$$\mathbb{P}\left(\left|\frac{\max_{i \in [p]} \epsilon_p(i)}{u_p} - 1\right| > \delta_p\right) \to 0, \quad \text{as } p \to \infty.$$

To this end, let

$$\xi_p := \frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i), \tag{5.1}$$

where $\mathscr{E} = \left\{\epsilon_p(i), i \in [p]\right\}$ is a URS Gaussian array with standard Normal marginals. Observe that

$$\begin{aligned} \mathbb{P}(|\xi_p - 1| > \delta_p) &= \mathbb{P}(\xi_p > 1 + \delta_p) + \mathbb{P}(\xi_p < 1 - \delta_p) \\ &=: \mathrm{I}(\delta_p) + \mathrm{II}(\delta_p). \end{aligned} \tag{5.2}$$

Thus, to obtain the desired rate we need to recover a bound on the rate of $\mathrm{I}(\delta_p)$ and $\mathrm{II}(\delta_p)$. Note that in our endeavor to secure upper bound on the term $\mathrm{II}(\delta_p)$ we will use the expectation of ξ_p. The integrability of ξ_p is ensured by Appendix A.2 of Chatterjee (2014), or Pickands (1968) in conjunction with Eq. 1.3.

Term I(δ_p) In the following proposition, we find an upper bound on the rate of δ_p in $\mathrm{I}(\delta_p)$ of Eq. 5.2. Interestingly, the following result does not involve the dependence structure of the array $\mathscr{E}$.

Proposition 5.1 *Let $\mathcal{E} = \{\epsilon_p(i), \ i \in [p]\}$ be an arbitrary Gaussian triangular array, where the marginal distributions are standard Normal and let ξ_p be defined as in Eq. 5.1. If $\delta_p \to 0$ is a positive sequence such that*

$$\delta_p \gg \frac{1}{\log(p)} \tag{5.3}$$

then, regardless of the dependence structure of $\mathcal{E}$, we have

$$\lim_{p \to \infty} \left(\delta_p^{-1} \mathbb{E}(\xi_p - 1)_+ \right) = 0, \tag{5.4}$$

and consequently $\mathbb{P}(\xi_p > 1 + \delta_p) \to 0$, as $p \to \infty$.

We need the following simple bound for the Mill's ratio (see also (1.2.2) or (2.1.1) in Adler and Taylor, 2009).

Lemma 5.1 *For all $u > 0$, we have*

$$1 - \frac{1}{1 \vee u^2} \leq \frac{\overline{\Phi}(u)}{\phi(u)/u} \leq 1,$$

where $\phi(u) = e^{-u^2/2}/\sqrt{2\pi}$ and $\overline{\Phi}(u) = \int_u^\infty \phi(x)dx$.

Proof We have

$$\frac{\overline{\Phi}(u)}{\phi(u)/u} = \frac{u}{\phi(u)} \int_u^\infty \phi(x)dx = u \int_u^\infty e^{-\frac{x^2 - u^2}{2}} dx$$
$$= u \int_0^\infty e^{-\frac{(z+u)^2 - u^2}{2}} dz = \int_0^\infty e^{-\frac{z^2}{2}} u e^{-uz} dz = \mathbb{E}[e^{-E^2/(2u^2)}],$$

where E is an exponentially distributed random variable with unit mean, and we used the change of variables $z := x - u$. Observing that $1 - x \leq e^{-x} \leq 1$, for all $x \geq 0$, we get

$$1 - \frac{E^2}{2u^2} \leq e^{-E^2/(2u^2)} \leq 1.$$

The result follows upon taking expectation and recalling that $\mathbb{E}[E^2] = 2$. $\qquad\square$

Proof (Proposition 5.1) Note first that Eq. 5.4 implies $\mathbb{P}(\xi_p > 1 + \delta_p) \to 0$. Indeed, this follows from the Markov inequality:

$$\mathbb{P}(\xi_p - 1 > \delta_p) = \mathbb{P}((\xi_p - 1)_+ > \delta_p) \leq \delta_p^{-1} \mathbb{E}(\xi_p - 1)_+.$$

Now, we focus on proving Eq. 5.4. We can write

$$\frac{1}{\delta_p} \mathbb{E}(\xi_p - 1)_+ = \frac{1}{\delta_p} \int_0^\infty \mathbb{P}(\xi_p - 1 > z)dz$$
$$= \int_0^\infty \mathbb{P}(\xi_p > 1 + \delta_p x)dx =: J(\delta_p), \tag{5.5}$$

where in the last integral we used the change of variables $z = \delta_p x$.

Recalling that $\xi_p = u_p^{-1} \max_{i \in [p]} \epsilon_p(i)$, by the union bound, for the last integrand we have that

$$\mathbb{P}(\xi_p > 1 + \delta_p x) \le p \overline{\Phi}(u_p(1 + \delta_p x)) = \frac{\overline{\Phi}(u_p(1 + \delta_p x))}{\overline{\Phi}(u_p)}. \tag{5.6}$$

By Lemma 5.1, we further obtain that

$$\frac{\overline{\Phi}(u_p(1 + \delta_p x))}{\overline{\Phi}(u_p)} \le \frac{1}{1 - 1/(1 \vee u_p^2)} \cdot \frac{\phi(u_p(1 + \delta_p x))}{(1 + \delta_p x)\phi(u_p)}$$

$$\le \frac{1}{1 - 1/(1 \vee u_p^2)} \exp\left\{ - \frac{u_p^2}{2}\left((1 + \delta_p x)^2 - 1\right)\right\}$$

$$\le B_p \exp\{-u_p^2 \delta_p x\}, \tag{5.7}$$

where $B_p := (1 - 1/(1 \vee u_p^2))^{-1} \to 1$, as $p \to \infty$, is a constant independent of $x \ge 0$ and in the last inequality we also used the simple bound $(1 + \delta_p x)^2 - 1 \ge 2\delta_p x$.

Condition (5.3) means that there is a sequence $\gamma(p)$ diverging to infinity slower than $\log(p)$ such that

$$\delta(p) = \frac{\gamma(p)}{\log(p)}.$$

Thus, by Relation (5.7) and the facts that $u_p^2 \sim 2\log(p)$ and $B_p \sim 1$, as $p \to \infty$, we obtain

$$\frac{\overline{\Phi}(u_p(1 + \delta_p x))}{\overline{\Phi}(u_p)} \le 2 \cdot e^{-2\gamma(p)x},$$

for all sufficiently large p. Since $\gamma(p) \to \infty$, Relation (5.6) and the Dominated Convergence Theorem applied to Eq. 5.5, implies

$$\lim_{p \to \infty} \mathrm{J}(\delta_p) \le \lim_{p \to \infty} \int_0^\infty 2e^{-2\gamma(p)x} dx = 0.$$

This completes the proof of Eq. 5.7. $\hfill\square$

Term II(δ_p) Handling term II of Eq. 5.2 is more involved and this is where the dependence structure of the array plays a role. We start by presenting a more careful reformulation of Lemma B.1 in Gao and Stoev (2020).

Lemma 5.2 *Let $(X_i)_{i=1}^p$ be p iid random variables with distribution F and density f, such that*

$$\mathbb{E}(X_i)_- \equiv \mathbb{E}(\max\{-X_i, 0\}) < \infty.$$

Denote the maximum of the X_i's as $M_p := \max_{i=1,\ldots,p} X_i$. Suppose that f is eventually decreasing, i.e., there exists a C_0 such that $f(x_1) \ge f(x_2)$ whenever $C_0 \le x_1 \le x_2$, then

$$\frac{\mathbb{E}M_p}{u_{p+1}} \ge (1 - F^p(C_0)) + \frac{\mathbb{E}[X_1 | X_1 < C_0]}{u_{p+1}} F^p(C_0),$$

where $u_{p+1} = F^\leftarrow(1 - 1/(p + 1))$.

Proof For the proof, refer to the proof of Lemma B.1 in Gao and Stoev (2020). □

Recall that a Gaussian triangular array $\mathscr{E} = \{\epsilon_p(j)\}_{j=1}^p$ with standard Normal marginals is said to be UDD if for every $\tau > 0$,

$$N_{\mathscr{E}}(\tau) := \sup_{p \in \mathbb{N}} \max_{i=1,\dots,p} \left| \{ \kappa \in [p] : \text{Cov}(\epsilon_p(i), \epsilon_p(\kappa)) > \tau \} \right| < \infty. \qquad (5.8)$$

That is, for every p and $i \in [p]$, there are at most $N_{\mathscr{E}}(\tau)$ indices κ, such that the covariance between $\epsilon_p(i)$ and $\epsilon_p(\kappa)$ exceeds τ.

The function $N_{\mathscr{E}}(\tau)$ encodes certain aspects of the dependence structure of the array $\mathscr{E}$. It will play a key role in the derivation of the upper bound on the rate of concentration of maxima. The next result is an extension of Proposition A.1 in Gao and Stoev (2020) tailored to our needs. For the benefit of the reader, we reproduce the key argument involving a packing construction and the Sudakov-Fernique bounds, which may be of independent interest.

Proposition 5.2 *For every UDD Gaussian array $\mathscr{E}$, and any subset $S_p \subseteq \{1, \dots, p\}$ with $q = |S_p|$, and $\tau \in (0, 1)$, we have that*

$$\mathbb{E}\left[\frac{\max\limits_{j \in S_p} \epsilon_p(j)}{u_q} \right] \geq \frac{u_{q/N_{\mathscr{E}}(\tau)+1}}{u_q} \sqrt{1-\tau} \left(1 - \frac{1}{2^{q/N_\epsilon(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_\epsilon(\tau)+1}} \cdot \frac{1}{2^{q/N_\epsilon(\tau)}} \right) \quad (5.9)$$

$$:= 1 - R_q, \qquad (5.10)$$

where $N_{\mathscr{E}}(\tau)$ is given in Eq. 5.8.

Remark 5.1 Note that without loss of generality we can assume $S_p = \{1, \dots, p\}$. We prove a slightly more general result, but the only application in this paper will be for $q = p$.

Proof Define the canonical (pseudo) metric on S_p,

$$d(i, j) = \sqrt{\mathbb{E}(\epsilon(i) - \epsilon(j))^2}.$$

This metric takes values between 0 and 2, since $\epsilon_p(i)$, $i = 1, \dots, p$, have zero means and unit variances. Fix $\tau \in (0, 1)$, take $\gamma = \sqrt{2(1-\tau)}$ and let Γ be a γ-packing of S_p. That is, let Γ be a subset of S_p, such that for any $i, j \in \Gamma$, $i \neq j$, we have $d(i, j) > \gamma$, i.e.,

$$d(i, j) = \sqrt{2\left(1 - \Sigma_p(i, j)\right)} \geq \gamma = \sqrt{2(1-\tau)},$$

or equivalently, $\Sigma_p(i, j) \leq \tau$. We claim that we can find a γ-packing Γ whose number of elements is at least

$$|\Gamma| \geq \frac{q}{N_{\mathscr{E}}(\tau)}. \qquad (5.11)$$

Indeed, Γ can be constructed iteratively as follows:

Step 1: Set $S_p^{(0)} := S_p$ and $\Gamma := \{j_1\}$, where $j_1 \in S_p^{(0)}$ is an arbitrary element. Set $k := 1$.

Step 2: Set $S_p^{(k)} := S_p^{(k-1)} \setminus B_\gamma(j_k)$, where

$$B_\gamma(j_k) := \{i \in S_p : d(i, j_k) < \gamma\}.$$

Step 3: If $S_p^{(k)} \neq \emptyset$, pick an arbitrary $j_{k+1} \in S_p^{(k)}$, set $\Gamma := \Gamma \cup \{j_{k+1}\}$, and $k := k + 1$, go to Step 2; otherwise stop.

By the definition of UDD, there are at most $N_{\mathscr{E}}(\tau)$ coordinates whose covariance with $\epsilon_p(j)$ exceed τ. Therefore, at each iteration, $|B_\gamma(j_k)| \leq N_{\mathscr{E}}(\tau)$, and hence

$$|S_p^{(k)}| \geq |S_p^{(k-1)}| - |B_\gamma(j_k)| \geq q - kN_{\mathscr{E}}(\tau).$$

The construction can continue for at least $q/N_{\mathscr{E}}(\tau)$ iterations, which implies Eq. 5.11.

Now, we define on this γ-packing Γ an independent Gaussian process $\{\eta(j)\}_{j \in \Gamma}$,

$$\eta(j) = \frac{\gamma}{\sqrt{2}} Z(j), \ j \in \Gamma,$$

where the $Z(j)$'s are iid standard Normal random variables. The increments of the new process are smaller than that of the original in the following sense,

$$\mathbb{E}(\eta(i) - \eta(j))^2 = \gamma^2 \leq d^2(i, j) = \mathbb{E}(\epsilon_p(i) - \epsilon_p(j))^2,$$

for all $i \neq j$, $i, j \in \Gamma$. Applying the Sudakov-Fernique inequality (see, e.g., Theorem 2.2.3 in Adler and Taylor 2009) to $\{\eta(j)\}_{j \in \Gamma}$ and $\{\epsilon_p(j)\}_{j \in \Gamma}$, we have

$$\mathbb{E}\left[\max_{j \in \Gamma}(\eta(j))\right] \leq \mathbb{E}\left[\max_{j \in \Gamma}(\epsilon_p(j))\right] \leq \mathbb{E}\left[\max_{j \in S_p}(\epsilon_p(j))\right].$$

This implies

$$\mathbb{E}\left[\frac{1}{u_q} \max_{j \in S_p} \epsilon_p(j)\right] \geq \mathbb{E}\left[\frac{1}{u_{|\Gamma|+1}} \max_{j \in \Gamma} \eta(j)\right] \cdot \frac{u_{|\Gamma|+1}}{u_q}$$

$$\geq \frac{u_{|\Gamma|+1}}{u_q} \cdot \sqrt{1 - \tau} \cdot \mathbb{E}\left[\frac{1}{u_{|\Gamma|+1}} \max_{j \in \Gamma} Z(j)\right].$$

Now, the application of Lemma 5.2 to the standard Normal distribution for $C_0 = 0$ entails that,

$$\frac{\mathbb{E}\left[\max_{j \in \Gamma} Z(j)\right]}{u_{|\Gamma|+1}} \geq 1 - \frac{1}{2^{|\Gamma|}} - \frac{\sqrt{2/\pi}}{u_{|\Gamma|+1}} \cdot \frac{1}{2^{|\Gamma|}}.$$

Since $|\Gamma| \geq q/N_{\mathscr{E}}(\tau)$ the desired lower bound in Eq. 5.9 is obtained. □

We are now interested in the rate at which the lower bound in Eq. 5.9 converges to 1. Equivalently, we desire to find the rate of decay of R_q. This rate is obtained in the following Lemma.

Lemma 5.3 *Let R_q, $\alpha(q)$ be defined as in* Eqs. 5.10 *and* 4.1 *respectively. Then*

$$R_q \asymp \alpha(q) + \tau(q) + 2^{-q^{1-\alpha(q)}}, \quad \text{as } q \to \infty. \tag{5.12}$$

Proof Note that by definition $R_q \to 0$, as $q \to \infty$. This implies that $R_q \sim \log(1 - R_q)$, as $q \to \infty$, so we just need the rate of

$$\log(1 - R_q) = \log\left(\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \cdot \sqrt{1 - \tau(q)} \cdot \left(1 - \frac{1}{2^{q/N_\epsilon(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_\epsilon(\tau)+1}} \cdot \frac{1}{2^{q/N_\epsilon(\tau)}}\right)\right)$$

$$= \log\left(\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q}\right) + \frac{1}{2}\log(1 - \tau(q))$$

$$+ \log\left(1 - \frac{1}{2^{q/N_\epsilon(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_\epsilon(\tau)+1}} \cdot \frac{1}{2^{q/N_\epsilon(\tau)}}\right).$$

Now, the facts that $\alpha(q) = \log(N_{\mathcal{E}}(\tau))/\log(q)$ and $u_q \sim \sqrt{2\log(q)}$ imply that

$$\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \sim \sqrt{\frac{2\log(1 + q^{1-\alpha(q)})}{2\log(q)}} \sim \sqrt{\frac{\log(q^{1-\alpha(q)})}{\log(q)}} = \sqrt{1 - \alpha(q)},$$

where we used the relation

$$q^{1-\alpha(q)} = e^{\log(q) - \log(N_{\mathcal{E}}(\tau))} = \frac{q}{N_{\mathcal{E}}(\tau)}. \tag{5.13}$$

However, since $\alpha(q) = \log(N_{\mathcal{E}}(\tau(q)))/\log(q) \to 0$ and $\tau(q) \to 0$, we have

$$\log(1 - \alpha(q)) = -\alpha(q) + o(\alpha(q)),$$
$$\log(1 - \tau(q)) = -\tau(q) + o(\tau(q)),$$

and by Eq. 5.13

$$\log\left(1 - \frac{1}{2^{q/N_\epsilon(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_\epsilon(\tau)+1}} \cdot \frac{1}{2^{q/N_\epsilon(\tau)}}\right)$$

$$= \log\left(1 - 2^{-q^{1-\alpha(q)}} - \frac{\sqrt{2/\pi}}{u_{q/N_\epsilon(\tau)+1}} \cdot 2^{-q^{1-\alpha(q)}}\right)$$

$$= 2^{-q^{1-\alpha(q)}} + o\left(2^{-q^{1-\alpha(q)}}\right).$$

As a result, we have

$$R_q \asymp \alpha(q) + \tau(q) + 2^{-q^{1-\alpha(q)}} + o\left(\max\left\{\alpha(q), \tau(q), 2^{-q^{1-\alpha(q)}}\right\}\right), \tag{5.14}$$

which completes the proof. □

Proof of Theorem 4.1 We are now in position to complete the proof of Theorem 4.1, which consists of a combination of the results that have already been established in Section 5.

Proof Recall the definition of ξ_p in Eq. 5.1 and that

$$\mathbb{P}(|\xi_p - 1| > \delta_p) = \mathrm{I}(\delta_p) + \mathrm{II}(\delta_p),$$

where $\mathrm{I}(\delta_p)$ and $\mathrm{II}(\delta_p)$ are defined as in Eq. 5.2. We shall show that both terms vanish.

Proposition 5.1, along with Eq. 4.2, imply that $\mathrm{I}(\delta_p) = \mathbb{P}(\xi_p > 1 + \delta_p) \to 0$, as $p \to \infty$. Observe that the term $\mathrm{I}(\delta_p) = \mathbb{P}(\xi_p > 1 + \delta_p)$ vanishes, regardless of the dependence structure of the array $\mathscr{E}$. The dependence plays a key role in the rate of the term $\mathrm{II}(\delta_p)$.

We now steer our focus towards term $\mathrm{II}(\delta_p)$. The Markov inequality yields

$$\mathrm{II}(\delta_p) = \mathbb{P}(\xi_p < 1 - \delta_p) \le \frac{\mathbb{E}(\xi_p - 1)_-}{\delta_p}.$$

Since $\mathbb{E}(\xi_p - 1)_- \le \mathbb{E}(\xi_p - 1)_+ + |\mathbb{E}(\xi_p - 1)|$, we have

$$\begin{aligned}
\mathrm{II}(\delta_p) &\le \frac{1}{\delta_p} \left(\mathbb{E}(\xi_p - 1)_+ + |\mathbb{E}(\xi_p - 1)| \right) \\
&= \frac{1}{\delta_p} \left(\mathbb{E}(\xi_p - 1)_+ + [\mathbb{E}(\xi_p - 1)]_+ + [\mathbb{E}(\xi_p - 1)]_- \right) \\
&\le \frac{1}{\delta_p} \left(2\mathbb{E}(\xi_p - 1)_+ + [\mathbb{E}(\xi_p - 1)]_- \right), \tag{5.15}
\end{aligned}$$

where the last inequality follows from the fact that $[\mathbb{E}(\xi_p - 1)]_+ \le \mathbb{E}(\xi_p - 1)_+$.

Proposition 5.1 and Eq. 4.2 imply that the term $\delta_p^{-1}\mathbb{E}(\xi_p - 1)_+$ in Eq. 5.15 vanishes. Moreover, Proposition 5.2 entails

$$[\mathbb{E}(\xi_p - 1)]_- = \max\{0, -\mathbb{E}(\xi_p - 1)\} \le |R_p|.$$

Thus, the term $\mathrm{II}(\delta_p)$ vanishes, provided that $R_p/\delta_p \to 0$. This follows, however, from Lemma 5.3 and Eq. 4.2, since for $\alpha(p) \to 0$, we have

$$\frac{1}{\log(p)} \gg 2^{-p^{1-\alpha(p)}}, \quad \text{as } p \to \infty$$

and the proof is complete. $\qquad\square$

Remark 5.2 After we completed and submitted this paper, we became aware of the important work of Tanguy (2015). According to their paper, in the stationary case, the upper bound of Theorem 4.1 above partially follows from their Theorem 3. However, our work is in the general setting of triangular arrays and does not require stationarity. The result in Theorem 5 of Tanguy (2015), could in principle, be used to derive bounds on rates of concentration of maxima for non-stationary arrays. This, however, requires verifying two technical conditions. Our approach, based on the UDD condition yields rates that can be explicitly related to the covariance structure of the array. The in-depth comparison of the two approaches merits an independent study beyond the scope of the present work.

References

Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Springer Science & Business Media (2009)

Arias-Castro, E., Chen, S.: Distribution-free multiple testing. Electron. J. Stat. **11**(1), 1983–2001 (2017)

Barndorff-Nielsen, O.: On the limit behaviour of extreme order statistics. Ann. Math. Stat. **34**(3), 992–1002 (1963)

Berman, S.M.: Limit theorems for the maximum term in stationary sequences. Ann. Math. Stat. **35**(2), 502–516 (1964)

Billingsley, P.: Measure and Probability. Wiley, New York (1995)

Billingsley, P.: Convergence of Probability Measures. Wiley, New York (2013)

Chatterjee, S.: Superconcentration and Related Topics, vol 15. Springer, Cham (2014)

Dasgupta, N., Spurrier, J.D.: A class of multivariate chi-square distributions with applications to comparison with a control. Commun. Stat. - Theory Methods **26**(7), 1559–1573 (1997). https://doi.org/10.1080/03610929708832000

Donoho, D., Jin, J.: Higher Criticism for detecting sparse heterogeneous mixtures. Ann. Stat. **32**(3), 962–994 (2004)

Gao, Z., Stoev, S.: Fundamental Limits of Exact Support Recovery in High Dimensions. Bernoulli **26**(4), 2605–2638 (2020)

Gnedenko, B.: Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. **44**(3), 423–453 (1943)

Halliwell, L.J.: The lognormal random multivariate. In: Casualty Actuarial Society E-Forum, p. 5 (2015)

Ingster, Y.I.: Minimax detection of a signal for l_n^p-balls. Math. Methods Stat. **7**(4), 401–428 (1998)

Ji, P., Jin, J.: UPS delivers optimal phase diagram in high-dimensional variable selection. Ann. Stat. **40**(1), 73–103 (2012)

Kinoshita, K., Resnick, S.I.: Convergence of scaled random samples in $\mathbb{R}^d$. Ann. Probab. **19**(4), 1640–1663 (1991)

Pickands, J. III.: Moment convergence of sample extremes. Ann. Math. Stat. **39**(3), 881–889 (1968)

Resnick, S.I., Tomkins, R.: Almost sure stability of maxima. J. Appl. Probab. **10**(2), 387–401 (1973)

Tanguy, K.: Some superconcentration inequalities for extrema of stationary Gaussian processes. Stat. Probab. Lett. **106**, 239–246 (2015)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.