**Inventiones Mathematicae manuscript No.**
(will be inserted by the editor)

# The dimension-free structure of nonhomogeneous random matrices

**Rafał Latała · Ramon van Handel · Pierre Youssef**

Received: date / Accepted: date

**Abstract** Let $X$ be a symmetric random matrix with independent but non-identically distributed centered Gaussian entries. We show that

$$\mathbf{E}\|X\|_{S_p} \asymp \mathbf{E}\left[\left(\sum_i \left(\sum_j X_{ij}^2\right)^{p/2}\right)^{1/p}\right]$$

for any $2 \leq p \leq \infty$, where $S_p$ denotes the $p$-Schatten class and the constants are universal. The right-hand side admits an explicit expression in terms of the variances of the matrix entries. This settles, in the case $p = \infty$, a conjecture of the first author, and provides a complete characterization of the class of infinite matrices with independent Gaussian entries that define bounded operators on $\ell_2$. Along the way, we obtain optimal dimension-free bounds on the moments $(\mathbf{E}\|X\|_{S_p}^p)^{1/p}$ that are of independent interest. We develop further extensions to non-symmetric matrices and to nonasymptotic moment and norm estimates for matrices with non-Gaussian entries that arise, for example, in the study of random graphs and in applied mathematics.

**Keywords** Random matrices · noncommutative probability · Schatten norms · nonasymptotic bounds

**Mathematics Subject Classification (2000)** 60B20 · 46B09 · 46L53 · 15B52

Rafał Latała
Institute of Mathematics, University of Warsaw, Banacha 2, 02-097, Warsaw, Poland
E-mail: rlatala@mimuw.edu.pl

Ramon van Handel
Fine Hall 207, Princeton University, Princeton, NJ 08544, USA
E-mail: rvan@princeton.edu

Pierre Youssef
Laboratoire de Probabilités, Statistique et Modélisation, Université Paris-Diderot, 5 rue Thomas Mann, 75205, Paris CEDEX 13, France
E-mail: youssef@lpsm.paris

## 1 Introduction and main results

The study of random matrices has long driven mathematical developments across a range of pure and applied mathematical disciplines. Initially motivated by questions arising in mathematical physics, classical random matrix theory (see, e.g., [16] for an introduction to this topic) is primarily concerned with random matrix models that possess a large degree of symmetry. For example, perhaps the most basic objects in this theory are matrices with independent and identically distributed entries, called Wigner matrices. Major advances on this subject were achieved in the past decade, resulting in an extremely detailed understanding of the fine-scale properties of the spectra of large Wigner and Wigner-like matrices.

There are however many situations in which classical random matrix models are of limited significance. For example, from a functional analytic perspective, one might naturally wish to view a random matrix as a random linear operator. One of the most basic questions one could ask in this context is under what conditions an (infinite) random matrix defines a bounded operator on $\ell_2$. Even this simple question appears, at present, to be almost entirely open. It is readily verified that such matrices could never have identically distributed entries; to obtain meaningful answers to such infinite-dimensional questions, it is therefore essential to consider nonhomogeneous random matrix models. In another context, many problems of applied mathematics, such as the analysis of random networks or numerical linear algebra, naturally give rise to structured random matrix models that are inherently nonhomogeneous. Such problems motivate the development of mathematical methods that can accurately capture the underlying structure.

The main approach to the study of general nonhomogeneous random matrices has been provided by variations on a classical result in noncommutative probability, the noncommutative Khintchine inequality of Lust-Piquard and Pisier [14, section 6]. We describe this inequality for concreteness in the setting of Gaussian symmetric matrices, though variants of it may be developed in much greater generality. Let $X$ be any symmetric random matrix with centered jointly Gaussian entries. It is readily verified that such a matrix can always be represented as $X = \sum_{i\geq 1} g_i A_i$, where $g_i$ are i.i.d. standard Gaussian variables and $A_i$ are given symmetric matrices. The noncommutative Khintchine inequality states that the moments of $X$ admit essentially the same estimates that would hold if $A_i$ were scalar quantities, that is, we have the following matrix analogue of the classical Khintchine inequalities:

$$\left\| \left( \sum_{i\geq 1} A_i^2 \right)^{\frac{1}{2}} \right\|_{S_p} \lesssim \mathbf{E}\|X\|_{S_p} \leq (\mathbf{E}\|X\|_{S_p}^p)^{1/p} \lesssim \sqrt{p} \left\| \left( \sum_{i\geq 1} A_i^2 \right)^{\frac{1}{2}} \right\|_{S_p}$$

for all $1 \leq p < \infty$, where we denote by $\|X\|_{S_p} := \mathrm{Tr}[|X|^p]^{1/p}$ the Schatten $p$-norm (that is, the $\ell_p$-norm of the singular values of $X$). This powerful estimate makes no assumption whatsoever on the covariance structure of the matrix entries; consequently, this result and its generalizations have had a major impact in noncommutative probability [14] as well as in applied mathematics [17].

Despite the significant power of the noncommutative Khintchine inequality, its conclusion remains in many ways unsatisfactory: both the upper and lower bounds become increasingly inaccurate for large $p$. For example, while this result characterizes, for fixed $p < \infty$, when an infinite random matrix is in the Schatten class $S_p$, it sheds little light on the question of which infinite random matrices define bounded operators on $\ell_2$ (the case $p = \infty$). In finite dimension $n$, one can still deduce useful quantitative bounds on the operator norm using that $\|X\|_{S_p} \asymp \|X\|_{S_\infty}$ for $p \sim \log n$. The resulting dimension-dependent bounds are notoriously inaccurate, however: they do not even capture correctly the norm of the most basic object in random matrix theory, the Gaussian Wigner matrix. These observations indicate that a detailed understanding of the spectral norms of nonhomogeneous random matrices will require far more precise information than is provided by the noncommutative Khintchine inequality. In the most general setting considered thus far, this aim remains out of reach. However, in this paper, we will settle these questions in what is perhaps the most natural case: that of nonhomogeneous random matrices with independent Gaussian (and some non-Gaussian) entries.

Let us now consider, therefore, any symmetric random matrix $X$ with independent centered Gaussian entries. Such a matrix may be represented as $X_{ij} = b_{ij}g_{ij}$, where $g_{ij}$ are i.i.d. standard Gaussian variables and $b_{ij} \geq 0$ are given scalars for $i \geq j$. In this case, the noncommutative Khintchine inequality reduces to

$$\left(\sum_i \left(\sum_j b_{ij}^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} \lesssim \mathbf{E}\|X\|_{S_p} \leq (\mathbf{E}\|X\|_{S_p}^p)^{1/p} \lesssim \sqrt{p}\left(\sum_i \left(\sum_j b_{ij}^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}}.$$

One of the main results of this paper is the following sharp form of this estimate.

**Theorem 1.1** *Let $X$ be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij}g_{ij}$, where $b_{ij} \geq 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i \geq j$. Then*

$$(\mathbf{E}\|X\|_{S_p}^p)^{1/p} \asymp \left(\sum_i \left(\sum_j b_{ij}^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} + \sqrt{p}\left(\sum_i \max_j b_{ij}^p\right)^{\frac{1}{p}}$$

*and*

$$\mathbf{E}\|X\|_{S_p} \asymp \left(\sum_i \left(\sum_j b_{ij}^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} + \max_{i \leq e^p} \max_j b_{ij}^* \sqrt{\log i} + \sqrt{p}\left(\sum_{i \geq e^p} \max_j b_{ij}^{*\,p}\right)^{\frac{1}{p}}$$

*for all $2 \leq p < \infty$, and*

$$\mathbf{E}\|X\|_{S_\infty} \asymp \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \sqrt{\log i}$$

*for $p = \infty$. Here the matrix $(b_{ij}^*)$ is obtained by permuting the rows and columns of the matrix $(b_{ij})$ such that $\max_j b_{1j}^* \geq \max_j b_{2j}^* \geq \cdots \geq \max_j b_{nj}^*$, and the constants in the estimates are universal (independent of $n, p, \{b_{ij}\}$).*

As a consequence we obtain, for example, a characterization of all infinite matrices with independent Gaussian entries that define bounded operators on $\ell_2$.

**Corollary 1.2** *Let $(X_{ij})_{i,j\in\mathbb{N}}$ be a symmetric infinite matrix with independent Gaussian entries $X_{ij} \sim N(a_{ij}, b_{ij}^2)$ for $i \geq j$. We have the following dichotomy:*

- *If*
$$\max_i \sum_j b_{ij}^2 < \infty, \qquad \max_{ij} b_{ij}^* \sqrt{\log i} < \infty, \qquad \|(a_{ij})\|_{S_\infty} < \infty,$$
*then $X$ defines a bounded operator on $\ell_2(\mathbb{N})$ a.s.*
- *Otherwise, $X$ is unbounded as an operator on $\ell_2(\mathbb{N})$ a.s.*

We will also develop a number of extensions of Theorem 1.1 to non-symmetric matrices and to matrices with non-Gaussian entries.

**Remark 1.3** While Theorem 1.1 is formulated for centered (zero mean) Gaussian matrices, the result is easily extended to noncentered matrices by noting that

$$\mathbf{E}[\|A + X\|_{S_p}^q]^{1/q} \asymp \|A\|_{S_p} + \mathbf{E}[\|X\|_{S_p}^q]^{1/q}$$

for any deterministic matrix $A$ and $p, q$ (see Remark 3.4 below). There is therefore no loss of generality in restricting attention to centered random matrices, as we will do for simplicity throughout the remainder of the paper.

At first sight, the statement of Theorem 1.1 may appear difficult to interpret. In fact, as we will presently explain, this result has a very simple probabilistic formulation that sheds significant light on the behavior of these matrices. Moreover, its proof will provide considerable insight into the structure of such matrices.

## 1.1 Probabilistic formulation

The questions investigated in this paper have their origin in a result of Seginer [15], which provides matching probabilistic upper and lower bounds on the operator norm of random matrices with independent and *identically distributed* entries with an arbitrary centered distribution: the expected operator norm of any such matrix is of the same order as the expectation of the maximal Euclidean norm of its rows and columns. The latter is a much simpler probabilistic quantity, which can be readily estimated in practice. The distribution of the entries is entirely irrelevant here: Seginer only uses that the distribution of the matrix is invariant under permutation of the entries, so that one may reduce the problem to a question about random matrices defined by random permutations. The latter can be investigated by combinatorial methods.

In view of such a remarkably general probabilistic principle, it is natural to ask whether the same conclusion also extends to nonhomogeneous matrix models that do not possess permutation symmetry. Unfortunately, this is not the case: as was already noted by Seginer, his result for i.i.d. matrices already fails to extend to the simplest examples of nonhomogeneous matrices with subgaussian entries. Surprisingly, however, no counterexamples could be found to the analogous question for *Gaussian* matrices. This led the first author to conjecture, about 15 years ago, that Seginer's conclusion might remain valid in general for nonhomogeneous random matrices with independent centered Gaussian entries (cf. [9]). Theorem 1.1 settles this conjecture in the affirmative: indeed, one may compute [7, Theorem 1.2]

$$\mathbf{E}\left[\max_i \sqrt{\sum_j X_{ij}^2}\right] \asymp \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \sqrt{\log i},$$

so that this quantity is always of the same order as $\mathbf{E}\|X\|_{S_\infty}$ by Theorem 1.1. Beside settling the conjecture, this observation furnishes the case $p = \infty$ of Theorem 1.1 with a natural probabilistic interpretation that is far from evident from the explicit expression. Let us note that, on the face of it, this conclusion is quite striking: it is trivial that the operator norm of a matrix must be large if the matrix possesses a row with large Euclidean norm; what we have shown is that for symmetric Gaussian matrices with independent centered entries, this is the *only* reason why the operator norm can be large, regardless of the variance pattern of the matrix entries. For some further discussion and a different probabilistic interpretation, see [7].

It turns out that the above probabilistic formulation can be developed in much greater generality. To this end, define the mixed norm

$$\|X\|_{\ell_p(\ell_2)} := \left(\sum_i \left(\sum_j X_{ij}^2\right)^{p/2}\right)^{1/p}$$

(the definition is extended in the obvious manner to the case $p = \infty$). As a consequence of Theorem 1.1, we will show that the distributions of the random variables $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$ are comparable in a strong sense.

**Corollary 1.4** *Under the assumptions of Theorem 1.1, we have*

$$\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \geq t] \leq \mathbf{P}[\|X\|_{S_p} \geq t] \leq C\,\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \geq t/C]$$

*for all $t \geq 0$ and $2 \leq p \leq \infty$, where $C$ is a universal constant. In particular,*

$$\mathbf{E}\Psi(\|X\|_{\ell_p(\ell_2)}) \leq \mathbf{E}\Psi(\|X\|_{S_p}) \leq \mathbf{E}\Psi(C\|X\|_{\ell_p(\ell_2)})$$

*for every increasing convex function $\Psi$ and $2 \leq p \leq \infty$.*

By taking $\Psi(x) = x^p$ or $\Psi(x) = x$, respectively, this result provides a tantalizing probabilistic interpretation of both explicit bounds that appear in Theorem 1.1: what we have shown is that for any symmetric Gaussian matrix with independent centered entries, the Schatten $p$-norm (that is, the $\ell_p$-norm of its eigenvalues) is of the same order as the $\ell_p$-norm of the Euclidean norms of its rows.

## 1.2 Main ideas of the proof

The proof of Theorem 1.1 builds on initial progress that was made in this direction in two earlier papers [2, 7].

In [2, Theorem 1.1], Bandeira and the second author proved the following bound on the operator norm of a Gaussian random matrix, which corresponds to the special case $p \sim \log n$ of the first inequality of Theorem 1.1 (cf. Theorem 3.7 below):

$$(\mathbf{E}\|X\|_{S_\infty}^{\log n})^{1/\log n} \asymp \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij} \sqrt{\log n}.$$

The proof relies on a comparison principle between the moments $\mathbf{E}[\mathrm{Tr}[X^{2p}]]$ of the $n$-dimensional matrix $X$ and the moments $\mathbf{E}[\mathrm{Tr}[Y^{2p}]]$ of an $r$-dimensional Wigner matrix $Y$, where $r$ depends only on the coefficients $b_{ij}$ and $p$. This elementary dimension-compression argument has proved to yield a very efficient mechanism to reduce questions about nonhomogeneous random matrices to analogous questions on homogeneous random matrices, which are already well understood.

By Jensen's inequality, the above result provides a dimension-dependent upper bound on the expected operator norm $\mathbf{E}\|X\|_{S_\infty}$ that turns out to be nearly sharp in many cases of interest. Nonetheless, the dimension-dependence makes this bound useless in infinite-dimensional situations, while our main aim is to obtain optimal dimension-free bounds. Unfortunately, this bound already yields the optimal result that could be obtained for the operator norm by the moment method. In order to surmount this obstacle, the second author developed in [7] an entirely different method to bound the operator norm of nonhomogeneous Gaussian matrices through the geometry of random processes. This approach made it possible to prove the following dimension-free bound, cf. [7, Theorem 1.4] and subsequent discussion:

$$\mathbf{E}\|X\|_{S_\infty} \lesssim \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \log i.$$

This result is strongly reminiscent of the sharp bound provided by Theorem 1.1 in the case $p = \infty$, but nonetheless falls short of this goal: to obtain the optimal estimate, the factor $\log i$ in the second term should be reduced to $\sqrt{\log i}$. While this may seem a small step away from the correct result, the suboptimal term appears to arise from a fundamental obstruction in this method of proof, cf. [7, section 4.3]. It is therefore unclear how this argument could be significantly improved.

In view of the two suboptimal bounds discussed above, it is natural to expect that the optimal result should arise by eliminating an inefficiency in the proof of one of these bounds. One of the main insights of this paper is that the missing ingredient for the proof of Theorem 1.1 (in the case $p = \infty$) lies in an entirely different place: to prove this result, we will exploit some strong structural information on the variance pattern of the underlying matrix. To this end, a key idea that we will develop is that, modulo a relabeling of the rows and columns, the random matrix $X$ can have bounded operator norm only if has a very specific form: it consists of a (nearly) block-diagonal "core", which is made up of blocks of controlled dimension in which the variances of the entries decay at a slow rate; and an off-diagonal remainder, in which the entry variances decay at a much faster rate (this structure will be made precise in Lemma 3.10 and is illustrated in Figure 3.1 below). Remarkably, it turns out that with this structure in hand, the suboptimal bounds described above already suffice to conclude the proof for $p = \infty$: the dimension-dependent bound optimally controls the diagonal blocks, which are low-dimensional by construction; while the dimension-free bound is sufficiently accurate to optimally control the remainder of the matrix. Beside enabling us to prove the result, this structural information provides considerable insight into what random matrices with bounded operator norm look like: they must be nearly block-diagonal, in precise sense.

For the general case where $2 \le p < \infty$, we will introduce a further decomposition of the matrix (illustrated in Figure 3.2): in one part of the matrix, the $S_p$-norm is of the same order as the $S_\infty$-norm, which was already addressed above; while in the remaining part, the expected $S_p$-norm is optimally captured by the corresponding moment bound. The main difficulty in this part of the proof is to obtain optimal bounds on the moments $(\mathbf{E}\|X\|_{S_p}^p)^{1/p}$ for $p \ll \log n$, which do not follow from the simple argument of [2]. While the proof is based on the same dimension-compression idea as in [2], the optimal implementation of this method is significantly more challenging and requires us to develop certain combinatorial "Brascamp-Lieb-type" inequalities that may be of independent interest (see section 2.2 below).

Let us note that the proofs of the case $p = \infty$ of Theorem 1.1 and of Corollaries 1.2 and 1.4 can be read independently of the more technical combinatorial arguments needed to extend our results to general Schatten norms. The reader who is interested primarily in the operator norm may skip ahead directly to section 3.2.

### 1.3 Non-symmetric matrices and non-Gaussian entries

So far we have stated our results in the setting of symmetric matrices with independent centered Gaussian entries. However, neither symmetry nor Gaussianity is essential, and we will develop various extensions of our main results to more general situations.

The easiest to eliminate is the symmetry assumption, which plays little role in the proofs and is largely made for notational convenience. All our results extend readily to non-symmetric matrices, as will be shown in section 4.1.

The question of non-Gaussian entries is more subtle. As was explained above, an example of Seginer [15, Theorem 3.2] shows that the conclusion of Corollary 1.4 *cannot* hold even for subgaussian variables. Despite superficial similarities, Corollary 1.4 arises for an entirely different reason than Seginer's result for i.i.d. matrices: we cannot prove a general comparison principle between $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$; instead, we prove explicit upper and lower bounds on each of these quantities in the Gaussian case, and show that these coincide. Gaussian analysis plays a crucial role in the proofs of these bounds, which suggests that the Gaussian distribution is rather special. Nonetheless, we will show in section 4.2 that Corollary 1.4 extends to a large class of non-Gaussian random matrices. Roughly speaking, we expect that the equivalence between $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$ should hold rather generally for entry distributions whose tails are at least as heavy as that of the Gaussian distribution, but not when the tails are strictly lighter than the Gaussian distribution. We will not develop this statement in full generality, but rather demonstrate this phenomenon in a broad class of heavy-tailed distributions.

While our results cannot yield optimal two-sided bounds for matrices with light-tailed entries, the explicit bounds of Theorem 1.1 still yield very useful *upper* bounds in this case. Of particular interest in this setting is the fact that, while Theorem 1.1 is sharp only up to universal constants, the sharpest moment bounds obtained in this paper possess nearly optimal constants. We will exploit this fact in section 4.3 to obtain some very sharp bounds on the norms of non-homogeneous matrices with bounded entries. For example, we will prove the following result:

**Theorem 1.5** *Let $X$ be an $n \times n$ symmetric matrix with independent, centered, and uniformly bounded entries for $i \geq j$. Then we have for every $p \in \mathbb{N}$*

$$(\mathbf{E}\|X\|_{S_{2p}}^{2p})^{1/2p} \leq 2\Bigg(\sum_i \bigg(\sum_j \mathbf{E}[X_{ij}^2]\bigg)^p\Bigg)^{\frac{1}{2p}} + C\sqrt{p}\Bigg(\sum_{i,j} \|X_{ij}\|_\infty^{2p}\Bigg)^{\frac{1}{2p}},$$

*where $C$ is a universal constant.*

This improves a result proved in [2] for the special case $p \sim \log n$ under an additional symmetry assumption. We will spell out several variations on this result in section 4.3 in view of their considerable utility in applications. For example, we will show that Theorem 1.5 effortlessly yields the correct behavior of the spectral edge of the centered adjacency matrix of Erdős-Rényi random graphs, recovering a recent result of [3] by a simpler and more general method.

### 1.4 Overview and notation

The rest of this paper is organized as follows. In section 2, we obtain optimal bounds on the moments $(\mathbf{E}\|X\|_{S_p}^p)^{1/p}$ in the Gaussian symmetric case, proving the first part of Theorem 1.1. The second part of Theorem 1.1, concerning the expected norms $\mathbf{E}\|X\|_{S_p}$, is proved in section 3. Here we also prove Corollary 1.4, as well as the characterization of infinite Gaussian matrices that define bounded operators on $\ell_2$. Finally, in section 4, we develop various extensions of our main results to non-symmetric and non-Gaussian matrices.

The following notation and terminology will be used throughout the paper. The Schatten norm $\|X\|_{S_p}$ and mixed norm $\|X\|_{\ell_p(\ell_2)}$ were already defined above. We will sometimes denote the operator norm as $\|X\| := \|X\|_{S_\infty}$ for notational simplicity. All random matrices will be real unless otherwise noted (however, the extension of our main results to complex matrices is essentially trivial, cf. section 2.3 below). We write $a \lesssim b$ or $b \gtrsim a$ to denote that $a \leq Cb$ for a universal constant $C$ (which does not depend on any parameter of the problem unless explicitly noted otherwise). We will write $a \asymp b$ if $a \lesssim b$ and $b \lesssim a$. Finally, we recall that a random variable $Z$ is said to be $\sigma$-subgaussian if $\mathbf{E}[e^{tZ}] \leq e^{t^2\sigma^2/2}$ for all $t \in \mathbb{R}$.

## 2 Moment estimates

The main result of this section is the following theorem.

**Theorem 2.1** *Let $X$ be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij} g_{ij}$, where $b_{ij} \geq 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i \geq j$. Then for $p \in \mathbb{N}$*

$$(\mathbf{E}[\mathrm{Tr}[X^{2p}]])^{1/2p} \leq 2\Bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^p\Bigg)^{1/2p} + 5\sqrt{2p}\bigg(\sum_i \max_j b_{ij}^{2p}\bigg)^{1/2p}.$$

This result essentially contains the first part of Theorem 1.1; it remains to extend the conclusion to arbitrary (non-integer) $p$ and to prove a matching lower bound. To this end, we will establish the following corollary:

**Corollary 2.2** *Under the assumptions of Theorem 2.1, we have for all $2 \leq p < \infty$*

$$(\mathbf{E}\|X\|_{S_p}^p)^{1/p} \asymp \Bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^{p/2}\Bigg)^{1/p} + \sqrt{p}\bigg(\sum_{i,j} b_{ij}^p\bigg)^{1/p}.$$

Together with Remark 2.3 below, this concludes the first part of Theorem 1.1. The remainder of this section is devoted to the proofs of these results.

The generality of Corollary 2.2 comes at the expense of replacing the sharp constants of Theorem 2.1 by larger universal constants. Theorem 2.1 is therefore of independent interest, and will be put to good use in section 4.2 below.

**Remark 2.3** It may appear somewhat strange that the second term in the upper bound of Theorem 2.1 is smaller than the second term in the lower bound of Corollary 2.2. It is instructive to give a direct proof that the two bounds are nonetheless of the same order. To this end, we can estimate using Young's inequality

$$\sum_{i,j} b_{ij}^{2p} \leq \sum_i \bigg(\sum_j b_{ij}^2\bigg) \max_j b_{ij}^{2p-2} \leq \frac{a^p}{p} \sum_i \bigg(\sum_j b_{ij}^2\bigg)^p + \frac{p-1}{p a^{p/(p-1)}} \sum_i \max_j b_{ij}^{2p}$$

for any $a > 0$. Therefore, setting $a = e^{-2(p-1)}$ we readily obtain

$$\bigg(\sum_{i,j} b_{ij}^{2p}\bigg)^{1/2p} \lesssim e^{-p}\Bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^p\Bigg)^{1/2p} + \bigg(\sum_i \max_j b_{ij}^{2p}\bigg)^{1/2p}.$$

The equivalence of the bounds in Theorem 2.1 and Corollary 2.2 (up to the value of the universal constant) follows immediately from this estimate.

## 2.1 Proof of Theorem 2.1

Before we begin developing the proof of Theorem 2.1 in earnest, we make a simple observation: it suffices to prove the result under the assumption that the diagonal entries of the matrix vanish.

**Lemma 2.4** *Let $X$ be defined as in Theorem 2.1, let $D_{ij} := X_{ij} 1_{i=j}$ be its diagonal part, and denote by $\tilde{X} := X - D$ its off-diagonal part. Then*

$$(\mathbf{E}[\mathrm{Tr}[X^{2p}]])^{1/2p} \leq (\mathbf{E}[\mathrm{Tr}[\tilde{X}^{2p}]])^{1/2p} + \sqrt{2p}\bigg(\sum_i b_{ii}^{2p}\bigg)^{1/2p}.$$

*Proof* This follows immediately using the triangle inequality and the simple estimate $(\mathbf{E}[\mathrm{Tr}[D^{2p}]])^{1/2p} = (\sum_i \mathbf{E}[X_{ii}^{2p}])^{1/2p} \le \sqrt{2p}(\sum_i b_{ii}^{2p})^{1/2p}$. □

In view of Lemma 2.4, it suffices to prove the result of Theorem 2.1 under the assumption that $b_{ii} = 0$ for all $i$ (with constant 4 rather than 5 in the second term). While this observation is trivial, the elimination of the diagonal entries will be very helpful for the implementation of the combinatorial arguments that are used in the proof. *We therefore assume in the rest of this section that $b_{ii} = 0$ for all $i$.*

We now turn to the main part of the proof. Our starting point is the identity

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] = \sum_{u_1,\ldots,u_{2p}\in[n]} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1}\,\mathbf{E}[g_{u_1u_2}g_{u_2u_3}\cdots g_{u_{2p}u_1}].$$

We view $u_1 \to u_2 \to \cdots \to u_{2p} \to u_1$ as a cycle in the complete undirected graph on $n$ points. By symmetry of the Gaussian distribution, each distinct edge $\{u_k, u_{k+1}\}$ that appears in the cycle must be traversed an even number of times for a term in the sum to be nonzero. We will call cycles with this property *even*.

Following [2], the *shape* $\mathbf{s}(\mathbf{u})$ of a cycle $\mathbf{u} \in [n]^{2p}$ is defined by relabeling the vertices in order of appearance. For example, the cycle $2 \to 7 \to 9 \to 7 \to 8 \to 7 \to 2$ has shape $1 \to 2 \to 3 \to 2 \to 4 \to 2 \to 1$. We define

$$\mathcal{S}_{2p} := \{\mathbf{s}(\mathbf{u}) : \mathbf{u} \text{ is an even cycle of length } 2p\}.$$

We denote by $n_i(\mathbf{s})$ the number of distinct edges that are traversed exactly $i$ times by a cycle of shape $\mathbf{s}$, and by $m(\mathbf{s})$ the number of distinct vertices visited by $\mathbf{s}$. In terms of these quantities, we can rewrite the moments $\mathbf{E}[\mathrm{Tr}[X^{2p}]]$ as

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] = \sum_{\mathbf{s}\in\mathcal{S}_{2p}} \prod_{i\ge1} \mathbf{E}[g^{2i}]^{n_{2i}(\mathbf{s})} \sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1},$$

where $g \sim N(0,1)$. The key difficulty of the proof is to obtain the following estimate.

**Proposition 2.5** *For any $\mathbf{s} \in \mathcal{S}_{2p}$*

$$\sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1} \le \left(\sum_i\left(\sum_j b_{ij}^2\right)^p\right)^{\frac{m(\mathbf{s})-1}{p}} \left(\sum_i \max_j b_{ij}^{2p}\right)^{1-\frac{m(\mathbf{s})-1}{p}}.$$

Proposition 2.5 will be proved in section 2.2 below. With this result in hand, we can readily complete the proof of Theorem 2.1 as in [2].

*Proof (Proof of Theorem 2.1)* Define for simplicity

$$\sigma_p := \left(\sum_i\left(\sum_j b_{ij}^2\right)^p\right)^{1/2p}, \qquad \sigma_p^* := \left(\sum_i \max_j b_{ij}^{2p}\right)^{1/2p}.$$

By rescaling the matrix $X$, we may assume without loss of generality that $\sigma_p^* = 1$. Then Proposition 2.5 implies the estimate

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] \le \sum_{\mathbf{s}\in\mathcal{S}_{2p}} \sigma_p^{2(m(\mathbf{s})-1)} \prod_{i\ge1} \mathbf{E}[g^{2i}]^{n_{2i}(\mathbf{s})}.$$

On the other hand, let $Y$ be an $r \times r$ symmetric matrix whose entries $Y_{ij}$ are i.i.d. standard Gaussian variables for $i \geq j$. Then we have

$$\begin{aligned}\mathbf{E}[\mathrm{Tr}[Y^{2p}]] &= \sum_{\mathbf{s}\in\mathcal{S}_{2p}} \mathrm{card}\{\mathbf{u} \in [r]^{2p} : \mathbf{s}(\mathbf{u}) = \mathbf{s}\} \prod_{i\geq 1} \mathbf{E}[g^{2i}]^{n_{2i}(\mathbf{s})} \\ &= \sum_{\mathbf{s}\in\mathcal{S}_{2p}} \frac{r!}{(r-m(\mathbf{s}))!} \prod_{i\geq 1} \mathbf{E}[g^{2i}]^{n_{2i}(\mathbf{s})}\end{aligned}$$

provided $p < r$ (which ensures that $r \geq m(\mathbf{s})$, as $m(\mathbf{s}) \leq p+1$ for any even cycle). As $(r-1)!/(r-m)! \geq (r-m+1)^{m-1}$ and using $m(\mathbf{s}) \leq p+1$, we obtain

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] \leq \frac{1}{r}\mathbf{E}[\mathrm{Tr}[Y^{2p}]] \leq \mathbf{E}[\|Y\|^{2p}] \quad \text{for} \quad r = \lfloor \sigma_p^2 \rfloor + p + 1.$$

We now invoke a standard bound on the norm of Wigner matrices [2, Lemma 2.2]

$$(\mathbf{E}[\mathrm{Tr}[X^{2p}]])^{1/2p} \leq \mathbf{E}[\|Y\|^{2p}]^{1/2p} \leq 2\sqrt{r} + 2\sqrt{2p} \leq 2\sigma_p + 4\sqrt{2p}$$

to conclude the proof. □

**Remark 2.6** Instead of Proposition 2.5, the following much simpler estimate

$$\sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1} \leq n\left(\max_i \sum_j b_{ij}^2\right)^{m(\mathbf{s})-1}\left(\max_{ij} b_{ij}^2\right)^{p-m(\mathbf{s})+1}$$

was obtained in [2, Lemma 2.5]. Indeed, this estimate is an almost immediate consequence of the fact that every edge of an even cycle is traversed at least twice. However, when combined with the rest of the argument, this estimate gives rise to a dimension-dependent bound on the Schatten norms

$$(\mathbf{E}[\mathrm{Tr}[X^{2p}]])^{1/2p} \lesssim n^{1/2p}\left(\max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}\sqrt{p}\right).$$

The latter coincides with the optimal bound of Theorem 2.1 only when $p \gtrsim \log n$. As we will presently see, the case $p \ll \log n$ is much more delicate, and the key feature of Proposition 2.5 is that it gives the right dimension-free estimate.

## 2.2 Proof of Proposition 2.5

To complete the proof of Theorem 2.1, it remains to prove the key estimate of Proposition 2.5. Its innocent-looking statement suggests that it should arise as an application of Hölder's inequality, and this is in essence the case. Nonetheless, we do not know a short proof of this fact. The difficulty in proving this result is that the manner in which Hölder's inequality must be applied depends nontrivially on the topology of the shape $\mathbf{s}$, and it is unclear at the outset how to identify the relevant structure. To facilitate the analysis, it will be convenient to place ourselves in a somewhat more general framework.

Let $\mathcal{G}_m$ be the set of all undirected, connected graphs $G = ([m], E(G))$. For any edge $e = \{i, j\}$ of a graph $G \in \mathcal{G}_m$ and $\mathbf{v} = (v_i)_{i\in[m]} \in [n]^m$, we will write $v(e) := \{v_i, v_j\}$. The key quantity that we will investigate is the following.

**Definition 2.7** For any graph $G \in \mathcal{G}_m$, and any family $\mathbf{k} := (k_e)_{e\in E(G)}$ of labelings of its edges by positive values $k_e > 0$, we define

$$W^{\mathbf{k}}(G) := \sum_{\mathbf{v}\in[n]^m} \prod_{e\in E(G)} b_{v(e)}^{k_e}.$$

Definition 2.7 is more general than the quantities that appear in Proposition 2.5. Indeed, given any shape $\mathbf{s}$ of length $2p$ with $m(\mathbf{s}) = m$ distinct vertices, define a graph $G \in \mathcal{G}_m$ whose edges are given by $E(G) = \{\{s_1, s_2\}, \{s_2, s_3\}, \ldots, \{s_{2p}, s_1\}\}$, and let $k_e$ be the number of times the edge $e \in E(G)$ is traversed by $\mathbf{s}$. Then clearly

$$\sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} b_{u_1u_2}b_{u_2u_3}\cdots b_{u_{2p}u_1} = \sum_{v_1\neq v_2\neq\cdots\neq v_m} \prod_{e\in E(G)} b_{v(e)}^{k_e} \le W^{\mathbf{k}}(G).$$

The conclusion of Proposition 2.5 will therefore follow immediately from the following more general statement about the quantity $W^{\mathbf{k}}(G)$.

**Theorem 2.8** *For any $G \in \mathcal{G}_m$ and $\mathbf{k}$ so that $k_e \ge 2$ for every $e \in E(G)$, we have*

$$W^{\mathbf{k}}(G) \le \left(\sum_i \left(\sum_j b_{ij}^2\right)^{\frac{|\mathbf{k}|}{2}}\right)^{\frac{2(m-1)}{|\mathbf{k}|}} \left(\sum_i \max_j b_{ij}^{|\mathbf{k}|}\right)^{1-\frac{2(m-1)}{|\mathbf{k}|}},$$

*where we denote $|\mathbf{k}| := \sum_{e\in E(G)} k_e$.*

The remainder of this section is devoted to the proof of this result.

### *2.2.1 Reduction to trees*

As a first step towards the proof of Theorem 2.8, we will show that it suffices to prove the result in the case where $G$ is a tree; the special structure of trees will be heavily exploited in the rest of the proof.

In the sequel, we denote by $\mathcal{G}_m^{\text{tree}}$ the set of trees on $m$ vertices, that is, the set of graphs $G \in \mathcal{G}_m$ such that $\operatorname{card} E(G) = m - 1$. We will also denote by $G_I$ the subgraph induced by a graph $G \in \mathcal{G}_m$ on a subset $I \subseteq [m]$ of its vertices.

When $G \in \mathcal{G}_m^{\text{tree}}$, the quantity $W^{\mathbf{k}}(G)$ takes a particularly simple form. Indeed, suppose without loss of generality that the vertices $[m]$ are ordered so that $m$ is a leaf of $G$ and every $\ell < m$ is a leaf of the induced subtree $G_{[\ell]}$. Then the tree is rooted at vertex 1, and every vertex $\ell \ge 2$ has a unique parent vertex $i_\ell \le \ell - 1$. In particular, then $E(G) = \{\{1, 2\}, \{i_3, 3\}, \ldots, \{i_m, m\}\}$, so we can write

$$W^{\mathbf{k}}(G) = \sum_{v_1,\ldots,v_m\in[n]} b_{v_1v_2}^{k_2} b_{v_{i_3}v_3}^{k_3} \cdots b_{v_{i_m}v_m}^{k_m},$$

where we denoted $k_{\{i_\ell,\ell\}} =: k_\ell$ for simplicity. Conversely, for any $k_2, \ldots, k_m > 0$ and $i_\ell \le \ell - 1$, the expression on the right-hand side arises as $W^{\mathbf{k}}(G)$ for some $G \in \mathcal{G}_m^{\text{tree}}$ (indeed, one may generate any tree by starting at the root and repeatedly attaching a new vertex $\ell$ to a previously generated vertex $i_\ell$).

We now show that among all graphs $G \in \mathcal{G}_m$, the value of $W^{\mathbf{k}}(G)$ is maximized by trees. This will allow us to restrict attention to trees in the rest of the proof.

**Lemma 2.9** *For any $G \in \mathcal{G}_m$ and $\mathbf{k} = (k_e)_{e\in E(G)}$ with $k_e > 0$, there exist $k'_2, \ldots, k'_m > 0$ with $\min_i k'_i \geq \min_e k_e$ and $\sum_{i=2}^m k'_i = |\mathbf{k}|$ such that*

$$W^{\mathbf{k}}(G) \leq \max_{i_3,\ldots,i_m\in[m]:i_\ell\leq\ell-1} \sum_{v_1,\ldots,v_m\in[n]} b_{v_1v_2}^{k'_2} b_{v_{i_3}v_3}^{k'_3} \cdots b_{v_{i_m}v_m}^{k'_m} \leq \max_{G'\in\mathcal{G}_m^{\mathrm{tree}}} W^{\mathbf{k}'}(G').$$

*Proof* Let $T$ be a spanning tree of $G$, and assume that the vertices $[m]$ of $G$ are ordered so that $m$ is a leaf of $T$ and every $\ell < m$ is a leaf of $T_{[\ell]}$ (this entails no loss of generality, as this can always be accomplished by relabeling the vertices of $G$). The only point of this assumption is that it ensures that the subgraph $G_{[\ell]}$ is connected for every $\ell$. We now define the numbers $k'_2, \ldots, k'_m$ as

$$k'_\ell := \sum_{e=\{i,\ell\}\in E(G_{[\ell]})} k_e.$$

That is, $k'_\ell$ is the total weight of edges incident to $\ell$ in $G_{[\ell]}$. It is clear from this definition that $\min_\ell k'_\ell \geq \min_e k_e$ and $\sum_\ell k'_\ell = |\mathbf{k}|$.

The main part of the proof proceeds by induction. For the initial step, we begin by isolating the contribution of the vertex $m$ in the definition of $W^{\mathbf{k}}(G)$:

$$\begin{aligned} W^{\mathbf{k}}(G) &= \sum_{\mathbf{v}\in[n]^m} \Bigg(\prod_{e\in E(G_{[m-1]})} b_{v(e)}^{k_e}\Bigg)\Bigg(\prod_{\{i,m\}\in E(G)} b_{v_iv_m}^{k(i)}\Bigg) \\ &= \sum_{\mathbf{v}\in[n]^m} \prod_{\{i,m\}\in E(G)} \Bigg(b_{v_iv_m}^{k'_m} \prod_{e\in E(G_{[m-1]})} b_{v(e)}^{k_e}\Bigg)^{k(i)/k'_m}, \end{aligned}$$

where we denote $k_{\{i,m\}} =: k(i)$ and we observe that $\sum_i k(i) = k'_m$. Therefore

$$\begin{aligned} W^{\mathbf{k}}(G) &\leq \prod_{\{i,m\}\in E(G)} \Bigg(\sum_{\mathbf{v}\in[n]^m} b_{v_iv_m}^{k'_m} \prod_{e\in E(G_{[m-1]})} b_{v(e)}^{k_e}\Bigg)^{k(i)/k'_m} \\ &\leq \max_{i_m\leq m-1} \sum_{\mathbf{v}\in[n]^m} b_{v_{i_m}v_m}^{k'_m} \prod_{e\in E(G_{[m-1]})} b_{v(e)}^{k_e} \end{aligned}$$

by Hölder's inequality.

The argument for the inductive step proceeds along very similar lines. Suppose we have shown, for some $4 \leq r \leq m$, the induction hypothesis

$$W^{\mathbf{k}}(G) \leq \max_{i_r,\ldots,i_m:i_\ell\leq\ell-1} \sum_{\mathbf{v}\in[n]^m} b_{v_{i_r}v_r}^{k'_r} \cdots b_{v_{i_m}v_m}^{k'_m} \prod_{e\in E(G_{[r-1]})} b_{v(e)}^{k_e}.$$

By our assumption on the ordering of the vertices, $G_{[r-1]}$ is connected. In particular, this means that there exists at least one edge between vertex $r-1$ and $[r-2]$. Isolating these edges in the same manner as above yields

$$\begin{aligned} &\sum_{\mathbf{v}\in[n]^m} b_{v_{i_r}v_r}^{k'_r} \cdots b_{v_{i_m}v_m}^{k'_m} \prod_{e\in E(G_{[r-1]})} b_{v(e)}^{k_e} = \\ &\sum_{\mathbf{v}\in[n]^m} \prod_{\{i,r-1\}\in E(G_{[r-1]})} \Bigg(b_{v_iv_{r-1}}^{k'_{r-1}} b_{v_{i_r}v_r}^{k'_r} \cdots b_{v_{i_m}v_m}^{k'_m} \prod_{e\in E(G_{[r-2]})} b_{v(e)}^{k_e}\Bigg)^{k(i)/k'_{r-1}}, \end{aligned}$$

where we now redefined $k_{\{i,r-1\}} =: k(i)$. Applying Hölder's inequality once more establishes the validity of the induction hypothesis for $r \leftarrow r-1$.

The above induction guarantees the validity of the induction hypothesis for $r = 3$. This completes the proof, however, as $G_{[2]}$ contains the single edge $\{1,2\}$. □

### 2.2.2 *Iterative pruning*

By virtue of Lemma 2.9, we have now reduced the proof of Theorem 2.8 to the special case where the graph $G$ is a tree. To complete the proof, we will iteratively apply Hölder's inequality to the leaves of the tree. Unlike in the proof of Lemma 2.9, however, it is important in the present case to keep track of the powers of the different terms generated by Hölder's inequality, which introduces additional complications. To facilitate the requisite bookkeeping, it will be convenient to consider a further generalization of the quantity $W^{\mathbf{k}}(G)$.

In the following, let us suppose that each edge $e \in E(G)$ may be endowed with an independent (symmetric) weight matrix $b^{(e)}_{ij}$, and define

$$W(G) := \sum_{\mathbf{v}\in[n]^m} \prod_{e\in E(G)} b^{(e)}_{v(e)}.$$

We will recover $W(G) = W^{\mathbf{k}}(G)$ by setting $b^{(e)}_{ij} = b^{k_e}_{ij}$.

**Lemma 2.10** *For any $G \in \mathcal{G}^{\mathrm{tree}}_m$ and $p_e \geq 1$ such that $\sum_{e\in E(G)} 1/p_e = 1$, we have*

$$W(G) \leq \prod_{e\in E(G)} \Bigg(\sum_i \bigg(\sum_j b^{(e)}_{ij}\bigg)^{p_e}\Bigg)^{1/p_e}.$$

This Hölder-type inequality is reminiscent of a special Brascamp-Lieb inequality (see, for example, [4] and the references therein), but involving mixed $\ell_p(\ell_1)$ norms. We do not know whether it follows from a more general principle.

*Proof* Throughout the proof we will assume without loss of generality that $m > 2$, as the conclusion is trivial in the case $m = 2$.

The proof again proceeds by induction. For the initial step, we begin by noting that any finite tree $G$ must have at least two leaves (that is, vertices that have exactly one neighbor). Suppose that vertices $\ell, \ell'$ are leaves of $G$. Then

$$W(G) = \sum_{\mathbf{v}\in[n]^I} \bigg(\sum_j b^{(e_\ell)}_{v_{i_\ell} j}\bigg)\bigg(\sum_j b^{(e_{\ell'})}_{v_{i_{\ell'}} j}\bigg) \prod_{e\in E(G_I)} b^{(e)}_{v(e)},$$

where $I = [m]\backslash\{\ell,\ell'\}$ and $e_\ell = \{i_\ell, \ell\}$, $e_{\ell'} = \{i_{\ell'}, \ell'\}$ are the unique edges connecting the leaves $\ell, \ell'$ to $I$ (here we used that as $m > 2$, the set $I$ is nonempty and $e_\ell \neq e_{\ell'}$). We can therefore estimate by Hölder's inequality

$$\begin{aligned} W(G) \leq {}& \Bigg[\sum_{\mathbf{v}\in[n]^I} \bigg(\sum_j b^{(e_\ell)}_{v_{i_\ell} j}\bigg)^{1+\frac{p_{e_\ell}}{p_{e_{\ell'}}}} \prod_{e\in E(G_I)} b^{(e)}_{v(e)}\Bigg]^{\frac{p_{e_{\ell'}}}{p_{e_\ell}+p_{e_{\ell'}}}} \times \\ & \Bigg[\sum_{\mathbf{v}\in[n]^I} \bigg(\sum_j b^{(e_{\ell'})}_{v_{i_{\ell'}} j}\bigg)^{1+\frac{p_{e_{\ell'}}}{p_{e_\ell}}} \prod_{e\in E(G_I)} b^{(e)}_{v(e)}\Bigg]^{\frac{p_{e_\ell}}{p_{e_\ell}+p_{e_{\ell'}}}}. \end{aligned}$$

Now observe the following properties of the right-hand side:

- The graph $G_I$ is again a tree, as we remove only leaves from $G$.
- Both sides of the inequality are 1-homogeneous in all the variables $b^{(e)}$.

The latter two properties will form the basis for the induction.

Let us now describe the induction step, which is again very similar. Suppose we have shown, for some $r > 1$, the induction hypothesis

$$W(G) \leq \prod_{s=1}^{S} \Bigg[ \sum_{\mathbf{v}\in[n]^{I_s}} \Bigg( \sum_j b^{(e_s)}_{v_{i_s} j} \Bigg)^{q_s} \prod_{e\in E(G_{I_s})} b^{(e)}_{v(e)} \Bigg]^{\frac{1}{\alpha_s}},$$

where $S < \infty$, $I_s \subseteq [m]$, $i_s \in I_s$, $e_s \in E(G)\backslash E(G_{I_s})$, and $\alpha_s, q_s \geq 1$ satisfy:

- $\operatorname{card} I_s = r$ and $G_{I_s}$ is a tree for every $s$.
- $q_s = \sum_{e\in E(G)\backslash E(G_{I_s})} p_{e_s}/p_e$.
- The right-hand side is 1-homogeneous in all the variables $b^{(e)}$, $e \in E(G)$.

We aim to show that the induction hypothesis remains valid for $r \leftarrow r - 1$.

Consider a single term $s$ on the right-hand side. As $G_{I_s}$ is a tree, it must have at least two leaves; in particular, there is a vertex $\ell \neq i_s$ that is a leaf of $G_{I_s}$. Thus

$$\sum_{\mathbf{v}\in[n]^{I_s}} \Bigg( \sum_j b^{(e_s)}_{v_{i_s} j} \Bigg)^{q_s} \prod_{e\in E(G_{I_s})} b^{(e)}_{v(e)} = \\ \sum_{\mathbf{v}\in[n]^{I'_s}} \Bigg( \sum_j b^{(e_s)}_{v_{i_s} j} \Bigg)^{q_s} \Bigg( \sum_j b^{(e')}_{v_{i'} j} \Bigg) \prod_{e\in E(G_{I'_s})} b^{(e)}_{v(e)},$$

where $I'_s = I_s\backslash\{\ell\}$ and $i' \in I'_s$ is the unique vertex such that $e' = \{i', \ell\} \in E(G_{I_s})$. Applying Hölder's inequality readily yields

$$\sum_{\mathbf{v}\in[n]^{I_s}} \Bigg( \sum_j b^{(e_s)}_{v_{i_s} j} \Bigg)^{q_s} \prod_{e\in E(G_{I_s})} b^{(e)}_{v(e)} \leq \Bigg[ \sum_{\mathbf{v}\in[n]^{I'_s}} \Bigg( \sum_j b^{(e_s)}_{v_{i_s} j} \Bigg)^{q'_s} \prod_{e\in E(G_{I'_s})} b^{(e)}_{v(e)} \Bigg]^{\frac{q_s}{q'_s}} \\ \times \Bigg[ \sum_{\mathbf{v}\in[n]^{I'_s}} \Bigg( \sum_j b^{(e')}_{v_{i'} j} \Bigg)^{q'} \prod_{e\in E(G_{I'_s})} b^{(e)}_{v(e)} \Bigg]^{\frac{1}{q'}},$$

where $q'_s = \sum_{e\in E(G)\backslash E(G_{I'_s})} p_{e_s}/p_e$ and $q' = \sum_{e\in E(G)\backslash E(G_{I'_s})} p_{e'}/p_e$ (here we used $q_s/q'_s + 1/q' = 1$). Observe in particular that by construction, both sides in this inequality have the same degree of homogeneity in each variable $b^{(e)}$.

We have now shown how to bound a single term $s$ in the induction hypothesis. To conclude the induction argument, we replace every term in the induction hypothesis by the upper bound obtained by this procedure. We claim that the resulting bound again satisfies the induction hypothesis with $r \leftarrow r - 1$, concluding the induction step. Indeed, by construction, each set $I'_s$ that appears in the new bound satisfies $\operatorname{card} I'_s = r - 1$ and $G_{I'_s}$ is a tree (it was obtained from $G_{I_s}$ by removing a leaf). Moreover, by construction each term has the correct power $q'_s$. Finally, as each term in the induction hypothesis has been replaced by a term with the same homogeneity in each variable $b^{(e)}$, it follows immediately that the new bound is still 1-homogeneous. This concludes the proof of the induction step.

The above induction guarantees the validity of the induction hypothesis for $r = 1$, that is, we have proved the following bound:

$$W(G) \le \prod_{s=1}^{S} \left[ \sum_i \left( \sum_j b_{ij}^{(e_s)} \right)^{p_{e_s}} \right]^{\frac{1}{\alpha_s}} = \prod_{e \in E(G)} \left[ \sum_i \left( \sum_j b_{ij}^{(e)} \right)^{p_e} \right]^{\frac{1}{\alpha_e}},$$

where we defined $1/\alpha_e = \sum_{s:e_s=e} 1/\alpha_{e_s}$. Moreover, the induction argument guarantees that the right-hand side is 1-homogeneous in all variables $b^{(e)}$, $e \in E(G)$. It must therefore necessarily be the case that $\alpha_e = p_e$, concluding the proof. □

Combining Lemmas 2.9 and 2.10 with the assumption of Theorem 2.8 yields:

**Corollary 2.11** *For any $G \in \mathcal{G}_m$ and $\mathbf{k} = (k_e)_{e \in E(G)}$ such that $k_e \ge 2$ for every $e \in E(G)$, there exist $k'_2, \ldots, k'_m \ge 2$ such that $\sum_{\ell=2}^m k'_\ell = |\mathbf{k}|$ and*

$$W^{\mathbf{k}}(G) \le \prod_{\ell=2}^{m} \left( \sum_i \left( \sum_j b_{ij}^{k'_\ell} \right)^{\frac{|\mathbf{k}|}{k'_\ell}} \right)^{\frac{k'_\ell}{|\mathbf{k}|}}.$$

We can now conclude the proof of Theorem 2.8.

*Proof (Proof of Theorem 2.8)* For any $2 \le k \le K$, note that

$$\begin{aligned} \sum_i \left( \sum_j b_{ij}^k \right)^{K/k} &\le \sum_i \left( \sum_j b_{ij}^2 \right)^{K/k} \max_j b_{ij}^{K(k-2)/k} \\ &\le \left( \sum_i \left( \sum_j b_{ij}^2 \right)^{K/2} \right)^{2/k} \left( \sum_i \max_j b_{ij}^K \right)^{(k-2)/k}, \end{aligned}$$

where the second inequality used Hölder. The conclusion of Theorem 2.8 follows by applying this estimate to every term of Corollary 2.11 (with $k = k'_\ell$, $K = |\mathbf{k}|$). □

## 2.3 Proof of Corollary 2.2

There are two separate issues that must be addressed in the proof of Corollary 2.2. The first is to prove a lower bound on $\mathbf{E}\|X\|_{S_p}^p$, which is elementary. The second is to extend the upper bound of Theorem 2.1 to non-integer values of $p$, which will be accomplished by complex interpolation.

We begin with the lower bound. We will need the following deterministic fact.

**Lemma 2.12** *For any (not necessarily symmetric) matrix $M$ and $2 \le p \le \infty$*

$$\|M\|_{S_p} \ge \|M\|_{\ell_p(\ell_2)} := \left( \sum_i \left( \sum_j M_{ij}^2 \right)^{p/2} \right)^{1/p}.$$

*Proof* Note that as $p \geq 2$, we can write

$$\|M\|_{S_p}^2 = \|MM^*\|_{S_{p/2}} = \sup_{\|Z\|_{S_{p/(p-2)}} \leq 1} \mathrm{Tr}[ZMM^*]$$
$$\geq \sup_{\|v\|_{p/(p-2)} \leq 1} \mathrm{Tr}[\mathrm{diag}(v)MM^*] = \left(\sum_i (MM^*)_{ii}^{p/2}\right)^{2/p}.$$

The result follows readily. □

*Proof (Proof of Corollary 2.2: lower bound)* We compute two distinct lower bounds. First, note that by Lemma 2.12 and Jensen's inequality,

$$(\mathbf{E}\|X\|_{S_p}^p)^{1/p} \geq \left(\sum_i \left(\sum_j \mathbf{E}X_{ij}^2\right)^{p/2}\right)^{1/p} = \left(\sum_i \left(\sum_j b_{ij}^2\right)^{p/2}\right)^{1/p}.$$

On the other hand, using again Lemma 2.12, we can estimate

$$(\mathbf{E}\|X\|_{S_p}^p)^{1/p} \geq \left(\sum_{i,j} \mathbf{E}|X_{ij}|^p\right)^{1/p} \gtrsim \sqrt{p}\left(\sum_{i,j} b_{ij}^p\right)^{1/p},$$

where we used $(\sum_j X_{ij}^2)^{p/2} \geq \sum_j |X_{ij}|^p$ and $\mathbb{E}[|g|^p]^{1/p} \asymp \sqrt{p}$ when $g$ is standard Gaussian. Averaging these two bounds concludes the proof of the lower bound. □

In the rest of this section, we will use standard facts and definitions from the theory of complex interpolation that can be found, for example, in [13, Chapter 8].

In order to apply complex interpolation, it will be most convenient to work with non-symmetric matrices rather than symmetric ones. Our basic object of study will be defined as follows. Let $(\tilde{g}_{ij})_{i,j\in[n]}$ be i.i.d. (real) standard Gaussian variables, and define the linear mapping $T : \mathbb{C}^{n\times n} \to \bigcap_p L^p(\Omega; \mathbb{C}^{n\times n})$ as

$$T((a_{ij})_{i,j\in[n]}) = (a_{ij}\tilde{g}_{ij})_{i,j\in[n]}.$$

That is, $T$ maps the complex coefficients $(a_{ij})$ to the non-symmetric complex random matrix with entries $(a_{ij}\tilde{g}_{ij})$. From Theorem 2.1, we deduce the following (here we define the random matrix norm $\|X\|_{L^p(S_p)} := (\mathbf{E}\|X\|_{S_p}^p)^{1/p}$).

**Lemma 2.13** *We have for all $p \in \mathbb{N}$*

$$\|T((a_{ij}))\|_{L^{2p}(S_{2p})} \lesssim \|(a_{ij})\|_{\ell_{2p}(\ell_2)} + \|(a_{ji})\|_{\ell_{2p}(\ell_2)} + \sqrt{p}\,\|(a_{ij})\|_{\ell_{2p}}.$$

*Proof* By writing $T((a_{ij})) = (\mathrm{Re}\, a_{ij}\tilde{g}_{ij}) + i(\mathrm{Im}\, a_{ij}\tilde{g}_{ij})$ and applying the triangle inequality, it evidently suffices to prove the claim for the case where all $a_{ij}$ are real. To this end, form the $2n \times 2n$ symmetric matrix

$$X = \begin{pmatrix} 0 & (a_{ij}\tilde{g}_{ij}) \\ (a_{ji}\tilde{g}_{ji}) & 0 \end{pmatrix},$$

and note that $\|X\|_{S_{2p}}^{2p} = \|X^2\|_{S_p}^p = 2\|(a_{ij}\tilde{g}_{ij})\|_{S_{2p}}^{2p}$. The conclusion now follows readily by applying Theorem 2.1 to the real symmetric random matrix $X$. □

Observe that all three norms that appear on the right-hand side of Lemma 2.13 are Banach lattice norms, as they are monotone in $|a_{ij}|$ for every $i, j$. This enables us to apply a very convenient observation of [11, Theorem 2]: intersections of Banach lattices are well-behaved under complex interpolation, in the sense that if $B_i = (\mathbb{C}^m, \|\cdot\|_{B_i})$, $i = 0, 1, 2$ are finite-dimensional Banach lattices, then $\|x\|_{(B_0,B_1\cap B_2)_\theta} \le 2\|x\|_{(B_0,B_1)_\theta\cap(B_0,B_2)_\theta}$, where $\|x\|_{B\cap C} := \max(\|x\|_B, \|x\|_C)$. It remains to combine this observation with standard interpolation arguments.

*Proof (Proof of Corollary 2.2: upper bound)* Let $1 \le p < \infty$ be arbitrary. Define $q := \lceil p \rceil$, and $\theta \in (0, 1)$ by $(1 - \theta)/2 + \theta/2q = 1/2p$. By Lemma 2.13, we have

$$\|T((a_{ij}))\|_{L^2(S_2)} \lesssim \|(a_{ij})\|_{\ell_2},$$
$$\|T((a_{ij}))\|_{L^{2q}(S_{2q})} \lesssim \|(a_{ij})\|_{\ell_{2q}(\ell_2)} + \|(a_{ji})\|_{\ell_{2q}(\ell_2)} + \sqrt{p}\,\|(a_{ij})\|_{\ell_{2q}}.$$

We now recall that

$$(\ell_2, \ell_{2q}(\ell_2))_\theta = \ell_{2p}(\ell_2), \qquad (\ell_2, \ell_{2q})_\theta = \ell_{2p}, \qquad (L^2(S_2), L^{2q}(S_{2q}))_\theta = L^{2p}(S_{2p})$$

isometrically, see [13, Theorem 8.21 and (14.3)]. Thus the above-mentioned lattice property [11] and the fundamental theorem of interpolation [13, Theorem 8.8] yield

$$\|T((a_{ij}))\|_{L^{2p}(S_{2p})} \lesssim \|(a_{ij})\|_{\ell_{2p}(\ell_2)} + \|(a_{ji})\|_{\ell_{2p}(\ell_2)} + \sqrt{p}\,\|(a_{ij})\|_{\ell_{2p}}.$$

That is, we have shown that Lemma 2.13 extends to all (non-integer) $1 \le p < \infty$.

Finally, let the symmetric matrix $X$ be as in Theorem 2.1. To deduce the conclusion of Corollary 2.2, note that we can estimate by the triangle inequality

$$\|X\|_{L^p(S_p)} \le \|T((b_{ij}1_{i\ge j}))\|_{L^p(S_p)} + \|T((b_{ij}1_{i<j}))\|_{L^p(S_p)},$$

where we used that the entries above (or below) the diagonal of $X$ are independent. The conclusion now follows readily from the above estimate on $T$. □

## 3 Norm estimates

The main result of this section is the following theorem. When combined with Corollary 2.2 and Remark 2.3, this concludes the proof of Theorem 1.1.

**Theorem 3.1** *Let $X$ be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij}g_{ij}$, where $b_{ij} \ge 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i \ge j$. Then for $2 \le p \le \infty$*

$$\mathbf{E}\|X\|_{S_p} \asymp \mathbf{E}\left[\left(\sum_i \left(\sum_j X_{ij}^2\right)^{p/2}\right)^{1/p}\right]$$
$$\asymp \left(\sum_i \left(\sum_j b_{ij}^2\right)^{p/2}\right)^{1/p} + \max_{i\le e^p} \max_j b^*_{ij}\sqrt{\log i} + \sqrt{p}\left(\sum_{i\ge e^p} \max_j {b^*_{ij}}^p\right)^{1/p},$$

*where the matrix $(b^*_{ij})$ is obtained by permuting the rows and columns of the matrix $(b_{ij})$ such that* $\max_j b^*_{1j} \ge \max_j b^*_{2j} \ge \cdots \ge \max_j b^*_{nj}$.

**Remark 3.2** Theorem 3.1 is concerned with the regime $2 \le p \le \infty$. That its conclusion remains valid in the regime $1 \le p < 2$ follows already from the much more general noncommutative Khintchine inequality [14, section 6], when specialized to symmetric matrices with independent Gaussian entries (the formulation for non-symmetric matrices is a bit more subtle). In fact, the noncommutative Khintchine inequality is valid in the range $1 \le p < \infty$, but becomes increasingly suboptimal for large $p$. The key novelty of Theorem 3.1 is that it captures the precise behavior for $p \to \infty$ in the setting where the matrix has independent entries. Our result is already qualitatively new for $p = \infty$; it implies, for example, the characterization of bounded infinite-dimensional random matrices of Corollary 1.2 in the introduction. The proof of the latter result will be given at the end of this section.

**Remark 3.3** Theorem 3.1 proves that $\mathbf{E}\|X\|_{S_p} \asymp \mathbf{E}\|X\|_{\ell_p(\ell_2)}$. It is interesting to note that Corollary 2.2 admits a similar interpretation: one may show that its conclusion can be rewritten as $(\mathbf{E}\|X\|_{S_p}^p)^{1/p} \asymp (\mathbf{E}\|X\|_{\ell_p(\ell_2)}^p)^{1/p}$. These observations suggest that perhaps other moments of the random variables $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$ may also be comparable. In fact, we will show that the distributions of these random variables are comparable in a much stronger sense, as was stated in Corollary 1.4 in the introduction. This result will also be proved at the end of this section.

**Remark 3.4** Throughout the paper, we focus attention on centered Gaussian matrices $X$. There is however no loss of generality in doing so: a matrix with arbitrary mean can always be reduced to the centered case using that

$$\mathbf{E}[\|A + X\|_{S_p}^q]^{1/q} \asymp \|A\|_{S_p} + \mathbf{E}[\|X\|_{S_p}^q]^{1/q}$$

for any deterministic matrix $A$, centered random matrix $X$, and $q \ge 1$. Indeed, the upper bound is obvious by the triangle inequality. To show the lower bound, note that $\mathbf{E}[\|A+X\|_{S_p}^q]^{1/q} \ge \|A\|_{S_p}$ by Jensen's inequality (as $\mathbf{E}X = 0$), while $\mathbf{E}[\|A+X\|_{S_p}^q]^{1/q} \ge \mathbf{E}[\|X\|_{S_p}^q]^{1/q} - \|A\|_{S_p}$ by the reverse triangle inequality. Adding twice the first inequality to the second inequality gives a the desired lower bound.

## 3.1 The mixed norm

Before we turn to the main part of the proof of Theorem 3.1, we first establish the explicit expression given there in terms of the coefficients $b_{ij}$. This explicit expression will play an important role in the subsequent analysis of the random matrix. In the special case $p = \infty$, this result was proved in [7, Theorem 1.2]; we extend it here to any value of $2 \le p \le \infty$.

We will need the following elementary result.

**Lemma 3.5** *Let $G = (G_1, \ldots, G_n)$ be independent with $G_i \sim N(0, \sigma_i^2)$. Then*

$$\mathbf{E}\|G\|_{\ell_p} \asymp \max_{i \le e^p} \sigma_i^* \sqrt{\log(i+1)} + \sqrt{p}\left(\sum_{i \ge e^p} \sigma_i^{*p}\right)^{1/p}$$

*for $2 \le p \le \infty$, where $(\sigma_i^*)$ is the nonincreasing rearrangement of $(\sigma_i)$. The upper bound remains valid without independence and when $G_i$ is only $\sigma_i$-subgaussian.*

*Proof* By permutation invariance, we can assume in the following without loss of generality that $\sigma_i$ are positive and nonincreasing (so that $\sigma_i = \sigma_i^*$).

Let us begin with the upper bound. By the triangle inequality, we have $\mathbf{E}\|G\|_{\ell_p} \le \mathbf{E}\|G_{\le e^p}\|_{\ell_p} + \mathbf{E}\|G_{\ge e^p}\|_{\ell_p}$, where $G_{\le k} := (G_1, \dots, G_{\lfloor k \rfloor})$, $G_{\ge k} := (G_{\lceil k \rceil}, \dots, G_n)$. Now recall that for a vector $x \in \mathbb{R}^k$, we have $\|x\|_{\ell_p} \le k^{1/p}\|x\|_{\ell_\infty}$. Therefore

$$\mathbf{E}\|G_{\le e^p}\|_{\ell_p} \le e\,\mathbf{E}\|G_{\le e^p}\|_{\ell_\infty} \lesssim \max_{i \le e^p} \sigma_i \sqrt{\log(i+1)},$$

where the last inequality can be found in [7, Lemma 2.3]. On the other hand,

$$\mathbf{E}\|G_{\ge e^p}\|_{\ell_p} \le (\mathbf{E}\|G_{\ge e^p}\|_{\ell_p}^p)^{1/p} \lesssim \sqrt{p}\left(\sum_{i \ge e^p} \sigma_i^p\right)^{1/p},$$

where we used that $\mathbf{E}[|G_i|^p]^{1/p} \lesssim \sigma_i \sqrt{p}$ when $G_i$ is $\sigma_i$-subgaussian. This concludes the proof of the upper bound. Note that the only assumption that was used so far is that each $G_i$ is $\sigma_i$-subgaussian; no independence was assumed.

We now turn to the lower bound. In the sequel, we will make use of the stronger assumptions that $G_i \sim N(0, \sigma_i^2)$ and that $(G_i)$ are independent. First, note that

$$\max_{i \le e^p} \sigma_i \sqrt{\log(i+1)} \asymp \mathbf{E}\|G_{\le e^p}\|_{\ell_\infty} \le \mathbf{E}\|G\|_{\ell_p},$$

where the first inequality is given in [7, Lemma 2.4] and the second inequality is trivial. On the other hand, let us note that

$$\sqrt{p}\left(\sum_{i \ge e^p} \sigma_i^p\right)^{1/p} \asymp (\mathbf{E}\|G_{\ge e^p}\|_{\ell_p}^p)^{1/p} \lesssim \mathbf{E}\|G_{\ge e^p}\|_{\ell_p} + \max_{i \ge e^p} \sigma_i \sqrt{p},$$

where the first inequality follows as $\mathbf{E}[|G_i|^p]^{1/p} \asymp \sigma_i \sqrt{p}$ when $G_i \sim N(0, \sigma_i^2)$, and the second follows using the triangle inequality and that $\|G_{\ge e^p}\|_{\ell_p} - \mathbf{E}\|G_{\ge e^p}\|_{\ell_p}$ is $\max_{i \ge e^p} \sigma_i$-subgaussian by Gaussian concentration [5, Theorem 5.8]. But as we assumed that $\sigma_i$ are nonincreasing, we evidently have

$$\max_{i \ge e^p} \sigma_i \sqrt{p} \le \sigma_{\lfloor e^p \rfloor} \sqrt{p} \le \max_{i \le e^p} \sigma_i \sqrt{\log(i+1)} \lesssim \mathbf{E}\|G_{\le e^p}\|_{\ell_p}.$$

We can therefore easily conclude that

$$\sqrt{p}\left(\sum_{i \ge e^p} \sigma_i^p\right)^{1/p} \lesssim \mathbf{E}\|G\|_{\ell_p},$$

and the proof is completed by averaging the two lower bounds on $\mathbf{E}\|G\|_{\ell_p}$. □

We are now ready to prove the explicit bound given in Theorem 3.1.

**Corollary 3.6** *Under the assumptions of Theorem 3.1, we have for all* $2 \le p \le \infty$

$$\mathbf{E}\left[\left(\sum_i \left(\sum_j X_{ij}^2\right)^{p/2}\right)^{1/p}\right] \asymp$$
$$\left(\sum_i \left(\sum_j b_{ij}^2\right)^{p/2}\right)^{1/p} + \max_{i \le e^p} \max_j b_{ij}^* \sqrt{\log i} + \sqrt{p}\left(\sum_{i \ge e^p} \max_j {b_{ij}^*}^p\right)^{1/p}.$$

*Proof* Let us begin with the upper bound. Define the vector $Z = (Z_1, \ldots, Z_n)$ with $Z_i := (\sum_j X_{ij}^2)^{1/2}$. Then we can estimate using the triangle inequality

$$\mathbf{E}\bigg[\bigg(\sum_i \bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\bigg] = \mathbf{E}\|Z\|_{\ell_p} \le \|\mathbf{E}Z\|_{\ell_p} + \mathbf{E}\|Z - \mathbf{E}Z\|_{\ell_p}.$$

It follows easily from Jensen's inequality that

$$\|\mathbf{E}Z\|_{\ell_p} \le \bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^{p/2}\bigg)^{1/p}.$$

On the other hand, by Gaussian concentration [5, Theorem 5.8], the random variable $Z_i - \mathbf{E}Z_i$ is $\max_j b_{ij}$-subgaussian. We therefore obtain

$$\mathbf{E}\|Z - \mathbf{E}Z\|_{\ell_p} \lesssim \max_{i \le e^p} \max_j b_{ij}^* \sqrt{\log(i+1)} + \sqrt{p}\bigg(\sum_{i \ge e^p} \max_j {b_{ij}^*}^p\bigg)^{1/p}$$

by Lemma 3.5, which completes the proof of the upper bound. (In the final bound, we estimated $\log(i + 1) \le 1 + \log i$ for aesthetic reasons; this does not entail any loss provided we slightly increase the constant in front of the first two terms.)

To prove the lower bound, denote by $k_i$ the entry of the $i$th row of the matrix that has the largest variance, that is, $b_{ik_i} = \max_j b_{ij}$. Then we have

$$\begin{aligned}\mathbf{E}\bigg[\bigg(\sum_i \bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\bigg] &\ge \mathbf{E}\bigg[\bigg(\sum_i |X_{ik_i}|^p\bigg)^{1/p}\bigg] \\ &\gtrsim \max_{i \le e^p} \max_j b_{ij}^* \sqrt{\log i} + \sqrt{p}\bigg(\sum_{i \ge e^p} \max_j {b_{ij}^*}^p\bigg)^{1/p}\end{aligned}$$

using Lemma 3.5. To be precise, note that the random variables $X_{ik_i}$ and $X_{jk_j}$ are either independent or identically equal, the latter happening if $k_i = j$ and $k_j = i$. However, as each independent variable appears at most twice in the vector $(X_{ik_i})$, it is readily verified that the conclusion of Lemma 3.5 remains valid in this setting modulo a suitable modification of the universal constants.

On the other hand, we can lower bound by Jensen's inequality

$$\mathbf{E}\bigg[\bigg(\sum_i \bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\bigg] = \mathbf{E}\|Z\|_{\ell_p} \ge \|\mathbf{E}Z\|_{\ell_p}.$$

Now note that by the Gaussian Poincaré inequality [5, Theorem 3.20], we have

$$\sum_j b_{ij}^2 = \mathbf{E}[Z_i^2] = \mathbf{E}[Z_i]^2 + \mathrm{Var}(Z_i) \le \mathbf{E}[Z_i]^2 + \max_j b_{ij}^2 \lesssim \mathbf{E}[Z_i]^2.$$

We therefore obtain

$$\mathbf{E}\bigg[\bigg(\sum_i \bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\bigg] \gtrsim \mathbf{E}\bigg[\bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\bigg],$$

and the proof is concluded by averaging the two lower bounds. □

### 3.2 The operator norm

The aim of this section is to prove Theorem 3.1 in the case $p = \infty$. In fact, this turns out to be the most interesting case: in the next section, we will see that the proof of Theorem 3.1 for arbitrary $2 \le p \le \infty$ follows rather quickly by combining the case $p = \infty$ with Theorem 2.1. In the following, we will denote the operator norm as $\|X\| := \|X\|_{S_\infty}$ for notational simplicity.

Before we describe the main construction behind the proof, we must first recall two suboptimal bounds on $\mathbf{E}\|X\|$. First, we observe that a useful bound can already be deduced from Theorem 2.1; for the operator norm, this result was obtained (by a significantly simpler variant of the proof of Theorem 2.1) in [2].

**Theorem 3.7 ([2], Theorem 1.1)** *Let $X$ be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij}g_{ij}$, where $b_{ij} \ge 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i \ge j$. Then*

$$\mathbf{E}\|X\| \lesssim \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}\sqrt{\log n}.$$

*Proof* As $\mathbf{E}\|X\| \le (\mathbf{E}\|X\|_{S_{2p}}^{2p})^{1/2p}$, we can apply Theorem 2.1 for any $p$. We conclude by observing that $\|x\|_{\ell_p} \asymp \|x\|_{\ell_\infty}$ for $x \in \mathbb{R}^n$ if we choose $p \sim \log n$. □

The problem with this bound is that it is dimension-dependent, while the sharp result of Theorem 3.1 is inherently dimension-free. Unfortunately, as is indicated by Corollary 2.2, any bound on the $p$th moment with $p \sim \log n$ must necessarily depend on the dimension $n$. One therefore cannot hope to obtain a dimension-free bound by an improvement of the moment method. Instead, an entirely different approach was introduced in [7] to obtain a dimension-free bound on $\mathbf{E}\|X\|$ through the theory of Gaussian processes. We recall the following result without proof.

**Theorem 3.8 ([7], paragraph after Theorem 1.4)** *In the setting of Theorem 3.7*

$$\mathbf{E}\|X\| \lesssim \max_i \sqrt{\sum_j b_{ij}^2} + \max_{ij} b_{ij}^* \log i.$$

The advantage of Theorem 3.8 is that it is dimension-free, in a manner strongly reminiscent of the sharp result of Theorem 3.1. However, the result is suboptimal in a different sense, as its second term is too large ($\log i$ rather than $\sqrt{\log i}$).

We will presently show that the sharp bound of Theorem 3.1 in the case $p = \infty$ can be obtained by efficiently exploiting the two suboptimal bounds of Theorems 3.7 and 3.8. This will be achieved by decomposing the matrix $X$ into different pieces: dominant pieces of small dimension, which are controlled optimally by the dimension-dependent bound, and a small remainder of large dimension, which can be controlled by the dimension-free bound. This decomposition not only allows us to conclude the proof, but also provides significant insight into the structure of large random matrices with bounded operator norm.

We now proceed to develop the details of the construction. Fix a matrix $X$ as in Theorem 3.1, and define for the remainder of this subsection the quantities

$$a := \max_i \sqrt{\sum_j b_{ij}^2}, \qquad b := \max_{ij} b_{ij}^* \sqrt{\log i}.$$

Our aim is to prove the upper bound of Theorem 3.1 in the case $p = \infty$, that is:

**Theorem 3.9** $\mathbf{E}\|X\| \lesssim a + b$.

Theorem 3.9 completes the proof of Theorem 3.1 in the case $p = \infty$, as the corresponding lower bound follows trivially from Lemma 2.12 and Corollary 3.6.

At the heart of the proof lies the observation that the quantities $a$ and $b$ provide different types of control on the coefficients $b_{ij}$. On the one hand, by definition,

$$\operatorname{card}\left\{i : \max_j b_{ij} > \frac{b}{\sqrt{\log k}}\right\} < k \tag{B}$$

for every $k$. On the other hand, for every given $j$,

$$\operatorname{card}\left\{i : b_{ij} > \frac{a}{\sqrt{k}}\right\} < k \tag{A}$$

for every $k$. In other words, the maximal entry across columns $\max_j b_{ij}$ decays, when rearranged in decreasing order, as $b/\sqrt{\log i}$; while inside each given column $j$, the entries $b_{ij}$ decay, when rearranged in decreasing order, as $a/\sqrt{i}$. Of course, the ordering of entries in each column is different, so we cannot simultaneously rearrange all columns in decreasing order. Instead, we construct one rearrangement that benefits from both properties by alternating between (B) and (A).

We now describe the construction. We choose a permutation $(i_1, \ldots, i_n)$ of the indices $\{1, \ldots, n\}$ by a simple algorithm. After $k$ steps of the algorithm, we will have selected exactly $N_k := 2^{2^k}$ indices which we denote by $I_k := \{i_1, \ldots, i_{N_k}\}$.

**Step 1** Choose $N_1$ indices $i$ for which the quantity $\max_j b_{ij}$ is largest.

**Step k. (a)** Among the remaining indices, choose $N_{k-1}N_{k-2}$ indices $i$ that contain the $N_{k-2}$ largest entries $b_{ij}$ of each column $j \in I_{k-1}$.

**Step k. (b)** Among the remaining indices, choose $N_k - N_{k-1} - N_{k-1}N_{k-2}$ indices $i$ for which the quantity $\max_j b_{ij}$ is largest.

Note that this iteration is well defined as $N_k - N_{k-1} - N_{k-1}N_{k-2} > 1$ for all $k \geq 2$, and as $|I_k| = N_k$ for each $k$ by construction. The algorithm terminates when we have selected all $n$ indices. In the remainder of the proof, we will always work with the rearrangement of the indices constructed by this algorithm. We will therefore assume from now on, without loss of generality, that the algorithm selects the identity permutation $i_k = k$ (otherwise we may permute the rows and columns of the matrix such that this is the case, which does not alter any of the quantities of interest; this is just a relabeling of the indices of the matrix).

The construction we have just given ensures that we can control the magnitudes of the entries in different parts of the matrix.

**Lemma 3.10** *After rearranging as above, the following hold for all $k \geq 1$:*

*i.* $b_{ij} \lesssim b2^{-k/2}$ *when* $i \geq N_k$.
*ii.* $b_{ij} \leq a2^{-2^{k-2}}$ *when* $j \leq N_k$ *and* $i \geq N_k + N_k N_{k-1}$.

*By symmetry, the identical bounds hold for* $b_{ji}$.

*Proof* After $k$ steps of the algorithm, we have selected at least

$$N_1 + \sum_{s=2}^{k} (N_s - N_{s-1} - N_{s-1}N_{s-2}) \geq N_{k-1}$$

indices $i$ for which the quantity $\max_j b_{ij}$ is largest. Therefore, by (B), we have $b_{ij} \leq b/\sqrt{\log N_{k-1}}$ whenever $i \geq N_k$. This proves part *i*. On the other hand, after $k+1$ steps of the algorithm, the $N_{k-1}$ largest entries $b_{ij}$ of each of the first $N_k$ columns $j$ are contained in the first $N_k + N_k N_{k-1}$ rows $i$. Therefore, when $j \leq N_k$ and $i \geq N_k + N_k N_{k-1}$, we have $b_{ij} \leq a/\sqrt{N_{k-1}}$ by (A). This proves part *ii*. □

Lemma 3.10 provides a lot of information about the structure of the matrix $X$. Near the diagonal, the matrix consists of a sequence of blocks of dimension $\sim N_k$ whose entries are of order $\lesssim b2^{-k/2}$. On the other hand, away from the diagonal, the entries of the matrix decay much more rapidly at a rate $\lesssim a2^{-2^{k-2}}$. We can therefore decompose our matrix into a block-diagonal part and a small remainder. More precisely, let us partition the indices $(i, j) \in [1, n]^2$ of $X$ into three parts:

$$E_1 = [1, M_1]^2 \cup \bigcup_{k \geq 1} [N_{2k}, M_{2k+1}]^2, \qquad E_2 = \bigcup_{k \geq 1} [N_{2k-1}, M_{2k}]^2 \backslash E_1,$$

and

$$E_3 = [1, n]^2 \backslash (E_1 \cup E_2),$$

where we defined $M_k := N_k + N_k N_{k-1}$. This partition is illustrated in Figure 3.1.

We are now ready to complete the proof of Theorem 3.9.

*Proof (Proof of Theorem 3.9)* We decompose $X = U + V + W$ where

$$U_{ij} := X_{ij} 1_{(i,j) \in E_1}, \qquad V_{ij} := X_{ij} 1_{(i,j) \in E_2}, \qquad W_{ij} := X_{ij} 1_{(i,j) \in E_3}.$$

It suffices to bound the norms of each of these matrices.

**The norms of $U$ and $V$.** The crucial point is that $U$ and $V$ are block-diagonal matrices, so that their norm is the maximum of the norms of the blocks. Let us denote by $U_1, U_2, \ldots$ the (disjoint) diagonal blocks of $U$. Then $U_1$ is a square matrix of dimension $M_1$ and $U_{k+1}$ is a square matrix of dimension $M_{2k+1} - N_{2k} + 1 \leq N_{2k+2}$ for $k \geq 1$. Moreover, by Lemma 3.10, the maximal coefficient $b_{ij}$ in $U_k$ is $\lesssim b2^{-k}$ for $k > 1$, while we can trivially bound $b_{ij} \lesssim b$ for coefficients in $U_1$. Thus Theorem 3.7 and the Gaussian Poincaré inequality [5, Theorem 3.20] yield

$$\mathbf{E}\|U_k\| \lesssim a + b2^{-k} \sqrt{\log N_{2k+2}} \lesssim a + b, \qquad \mathrm{Var}(\|U_k\|)^{1/2} \lesssim b2^{-k}.$$

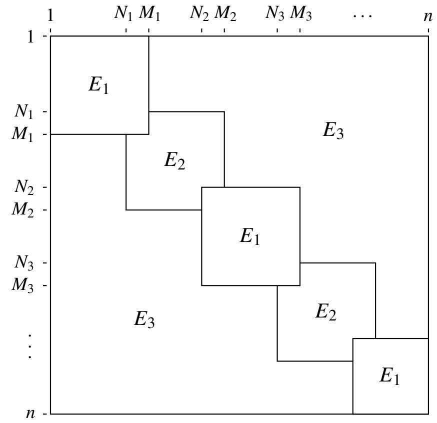


**Fig. 3.1** Illustration of the matrix decomposition used in the proof of Theorem 3.9. The figure is drawn for clarity on a log log scale.

We can therefore estimate

$$\begin{aligned}\mathbf{E}\|U\| = \mathbf{E}\max_k \|U_k\| &\le \max_k \mathbf{E}\|U_k\| + \mathbf{E}\max_k |\,\|U_k\| - \mathbf{E}\|U_k\|\,| \\ &\le \max_k \mathbf{E}\|U_k\| + \sum_k \mathbf{E}|\,\|U_k\| - \mathbf{E}\|U_k\|\,| \\ &\le \max_k \mathbf{E}\|U_k\| + \sum_k \mathrm{Var}(\|U_k\|)^{1/2} \lesssim a + b.\end{aligned}$$

The same bound on $\mathbf{E}\|V\|$ follows from the identical argument.

**The norm of $W$.** We now have to show that the remaining part of the matrix is small. To this end, let us verify that the following holds:

$$\max_j b_{ij} 1_{(i,j)\in E_3} \lesssim a 2^{-2^{k-2}} \quad \text{whenever } M_k \le i < M_{k+1}$$

holds for all $k \ge 1$. Indeed, it is readily read off from the definition of $E_3$ than when $M_k \le i < N_{k+1}$ and $(i, j) \in E_3$, then either $j \le N_k$ or $j \ge M_{k+1}$. In either case, Lemma 3.10 shows that $b_{ij} \lesssim a2^{-2^{k-2}}$. On the other hand, if $N_{k+1} \le i < M_{k+1}$, then either $j \le N_k$ or $j \ge M_{k+2}$. Again, Lemma 3.10 shows that $b_{ij} \lesssim a2^{-2^{k-2}}$. Combining these cases yields the claim. But we can now estimate by Theorem 3.8

$$\begin{aligned}\mathbf{E}\|W\| &\lesssim a + \max_{k\ge 1} \max_{M_k \le i < M_{k+1}} \log i \, \max_j b_{ij} 1_{(i,j)\in E_3} \\ &\lesssim a + \max_{k\ge 1} a 2^{-2^{k-2}} \log M_{k+1} \lesssim a,\end{aligned}$$

where we have trivially bounded $\max_{i<M_1} \log i \, \max_j b_{ij} 1_{(i,j)\in E_3} \lesssim a$. Combining the estimates for the norms of $U, V, W$ completes the proof. □

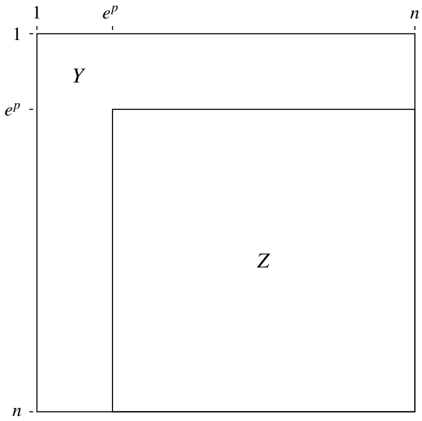


**Fig. 3.2** Matrix decomposition used in the proof of Theorem 3.1.

## 3.3 Proof of Theorem 3.1

Now that we have proved Theorem 3.1 in the case $p = \infty$, it remains to extend the conclusion to arbitrary $2 \le p \le \infty$. Perhaps somewhat surprisingly, this does not require significant additional work. The reason is already visible in the formula of Lemma 3.5: the behavior of the expected $\ell_p$ norm $\mathbf{E}\|G\|_{\ell_p}$ of a Gaussian vector $G$ interpolates between the sharp bound on the uniform norm $\mathbf{E}\|G\|_{\ell_\infty}$ and the sharp bound for the moments $(\mathbf{E}\|G\|_{\ell_p}^p)^{1/p}$. We will show that an analogous situation occurs for Schatten norms: we will bound $\mathbf{E}\|X\|_{S_p}$ by decomposing the random matrix $X$ into two parts, one of which is controlled by the sharp bound on the operator norm obtained in the previous section, and the other is controlled by the sharp moment bound provided by Corollary 2.2.

*Proof (Proof of Theorem 3.1)* The lower bound on $\mathbf{E}\|X\|_{S_p}$ follows trivially from Lemma 2.12, so it remains to prove the upper bound. We may assume without loss of generality that the rows and columns of $X$ have been permuted so that $\max_j b_{ij}$ is nonincreasing, that is, $b_{ij} = b^*_{ij}$. We begin by decomposing the matrix as

$$X = Y + Z, \qquad Z_{ij} := X_{ij}\mathbf{1}_{\min(i,j)\ge e^p}.$$

This decomposition is illustrated in Figure 3.2.

Consider first the matrix $Y$. Observe that $\mathrm{rank}(Y) \le 2e^p$, so we have $\|Y\|_{S_p} \le 2^{1/p}e\|Y\|_{S_\infty}$. We can therefore apply Theorem 3.9 and Corollary 3.6 to estimate

$$\mathbf{E}\|Y\|_{S_p} \lesssim \mathbf{E}\Bigg[\max_i \sqrt{\sum_j Y_{ij}^2}\Bigg] \le \mathbf{E}\Bigg[\Bigg(\sum_i \Bigg(\sum_j X_{ij}^2\Bigg)^{p/2}\Bigg)^{1/p}\Bigg].$$

We now consider the matrix $Z$. Corollary 2.2 and Remark 2.3 yield

$$\mathbf{E}\|Z\|_{S_p} \lesssim \Bigg(\sum_i \Bigg(\sum_j b_{ij}^2\Bigg)^{p/2}\Bigg)^{1/p} + \sqrt{p}\Bigg(\sum_{i\ge e^p} \max_j b_{ij}^p\Bigg)^{1/p}.$$

Therefore, by Corollary 3.6, we obtain once more

$$\mathbf{E}\|Z\|_{S_p} \lesssim \mathbf{E}\Bigg[\bigg(\sum_i\bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\Bigg].$$

It remains to apply the triangle inequality $\mathbf{E}\|X\|_{S_p} \le \mathbf{E}\|Y\|_{S_p} + \mathbf{E}\|Z\|_{S_p}$. □

### 3.4 Proof of Corollaries 1.2 and 1.4

The proof of Corollary 1.2 follows easily from Theorem 3.1 and standard arguments.

*Proof (Proof of Corollary 1.2)* It is an elementary fact [1, section II.26] that an infinite matrix $X$ defines a (necessarily unique) bounded operator on $\ell_2(\mathbb{N})$ if and only if $\sup_n \|X_{[n]}\|_{S_\infty} < \infty$, where we defined the finite submatrices $X_{[n]} := (X_{ij})_{i,j\le n}$.

Now note that by Theorem 3.1, Remark 3.4, and the monotone convergence theorem, we have $\mathbf{E}[\sup_n \|X_{[n]}\|_{S_\infty}] < \infty$ if and only if

$$\max_i \sum_j b_{ij}^2 < \infty, \qquad \max_{ij} b_{ij}^* \sqrt{\log i} < \infty, \qquad \|(a_{ij})\|_{S_\infty} < \infty.$$

Therefore, if the latter conditions hold $X$ defines a bounded operator a.s. Conversely, if any of these conditions fails, then $\mathbf{E}[\sup_n \|X_{[n]}\|_{S_\infty}] = \infty$. By a classical zero-one law for Gaussian measures [8], it follows that in fact $\sup_n \|X_{[n]}\|_{S_\infty} = \infty$ a.s. Thus in this case $X$ is unbounded as an operator on $\ell_2(\mathbb{N})$ a.s. □

We now turn to the proof of Corollary 1.4.

*Proof (Proof of Corollary 1.4)* We begin by proving the tail bounds. The lower bound is trivial by Lemma 2.12, so it suffices to consider the upper bound.

In the following we denote by $m$ the median of $\|X\|_{S_p}$. By the Gaussian isoperimetric theorem [5, Theorem 10.17], we can estimate for all $t \ge 0$

$$\mathbf{P}[\|X\|_{S_p} - m \ge t] \le \mathbf{P}\Big[\sqrt{2}\max_{ij} b_{ij}\, g \ge t\Big],$$

where $g \sim N(0,1)$. Moreover, as the median of a nonnegative random variable is bounded by twice its mean (this is a simple consequence of Markov's inequality), we have $m \le 2\mathbf{E}\|X\|_{S_p} \le K\,\mathbf{E}\|X\|_{\ell_p(\ell_2)}$ for a universal constant $K$ by Theorem 3.1.

We consider separately two cases. Suppose first that $t \ge 2m$. Then

$$\mathbf{P}[\|X\|_{S_p} \ge t] \le \mathbf{P}[\|X\|_{S_p} - m \ge t/2] \le \mathbf{P}\Big[2\sqrt{2}\max_{ij} b_{ij} g \ge t\Big] \le \mathbf{P}[2\sqrt{2}\|X\|_{\ell_p(\ell_2)} \ge t],$$

where we used that $\|X\|_{\ell_p(\ell_2)} \ge X_{ij} \sim b_{ij} g$ for any $i, j$.

Now consider the case $t \le 2m$. Using $m \le K\,\mathbf{E}\|X\|_{\ell_p(\ell_2)}$, we can estimate

$$\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \ge t/4K] \ge \mathbf{P}[\|X\|_{\ell_p(\ell_2)} \ge \tfrac{1}{2}\mathbf{E}\|X\|_{\ell_p(\ell_2)}] \ge \frac{1}{4}\frac{(\mathbf{E}\|X\|_{\ell_p(\ell_2)})^2}{\mathbf{E}\|X\|_{\ell_p(\ell_2)}^2}$$

by the Paley-Zygmund inequality. On the other hand, we can estimate

$$\mathrm{Var}(\|X\|_{\ell_p(\ell_2)}) \le \max_{ij} b_{ij}^2 = \tfrac{\pi}{2} \max_{ij} (\mathbf{E}|X_{ij}|)^2 \le \tfrac{\pi}{2} (\mathbf{E}\|X\|_{\ell_p(\ell_2)})^2$$

by the Gaussian Poincaré inequality [5, Theorem 3.20]. We have therefore shown that

$$\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \ge t/2K] \ge \tfrac{1}{2\pi+4} \ge \tfrac{1}{2\pi+4} \mathbf{P}[\|X\|_{S_p} \ge t]$$

for $t \le 2m$. Combining the above bounds yields the desired tail bound.

It remains to deduce the resulting bounds for convex functions. The lower bound is again trivial. To prove the upper bound, note first that

$$\mathbf{E}\Phi(\|X\|_{S_p}) = \Phi(0) + \int_0^\infty \Phi'(t)\, \mathbf{P}[\|X\|_{S_p} \ge t]\, dt$$

for any increasing function $\Phi$. We conclude immediately that

$$\mathbf{E}\Phi(\|X\|_{S_p}) \le C\, \mathbf{E}\Phi(C\|X\|_{\ell_p(\ell_2)})$$

for every nonnegative increasing function $\Phi$. As compared to the bound stated in Corollary 1.4, we have not assumed $\Phi$ is convex, but we have an additional constant $C$ in front of the expectation on the right-hand side. To eliminate it, let $\Psi$ be an increasing convex function, and choose $\Phi(t) = \Psi(Ct) - \Psi(0)$. By convexity, we have $\Phi(t) \ge C\, \Phi(t/C)$ for all $t \ge 0$. Substituting into the above bound yields

$$C\, \mathbf{E}\Phi(\|X\|_{S_p}/C) \le \mathbf{E}\Phi(\|X\|_{S_p}) \le C\, \mathbf{E}\Phi(C\|X\|_{\ell_p(\ell_2)}),$$

which yields upon rearranging

$$\mathbf{E}\Psi(\|X\|_{S_p}) \le \mathbf{E}\Psi(C^2\|X\|_{\ell_p(\ell_2)}).$$

Thus the claimed bound follows for a sufficiently large universal constant $C$. □

## 4 Extensions and complements

### 4.1 Non-symmetric matrices

For simplicity, we have restricted our attention so far to symmetric matrices $X$ with independent entries above the diagonal. However, the analogous results for non-symmetric matrices can be readily deduced in complete generality, modulo universal constants. Let us state, for example, the following analogue of Corollary 1.4 for non-symmetric matrices.

**Corollary 4.1** *Let $X$ be an $n \times m$ matrix with $X_{ij} = b_{ij} g_{ij}$, where $b_{ij} \ge 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i \le n$, $j \le m$. Then*

$$\mathbf{P}[\|X\|_{\ell_p(\ell_2)} + \|X^*\|_{\ell_p(\ell_2)} \ge 2t] \le \mathbf{P}[\|X\|_{S_p} \ge t] \le C\, \mathbf{P}[\|X\|_{\ell_p(\ell_2)} + \|X^*\|_{\ell_p(\ell_2)} \ge t/C]$$

*for all $t \ge 0$ and $2 \le p \le \infty$, where $C$ is a universal constant.*

*Proof* We use a standard device that already appeared in the proof of Lemma 2.13. Let $\tilde{X}$ be the $(n+m)\times(n+m)$ symmetric matrix defined by

$$\tilde{X} = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix},$$

and note that

$$\|\tilde{X}\|_{S_p}^p = \|\tilde{X}^2\|_{S_{p/2}}^{p/2} = 2\|X\|_{S_p}^p, \qquad \|\tilde{X}\|_{\ell_p(\ell_2)}^p = \|X\|_{\ell_p(\ell_2)}^p + \|X^*\|_{\ell_p(\ell_2)}^p.$$

The conclusion now follows readily by applying Corollary 1.4 to $\tilde{X}$. □

The only drawback of the approach of Corollary 4.1 is that it may not capture sharp constants. As most of the bounds in this paper are sharp only up to universal constants to begin with, this entails no further loss in our main results. The exception to this statement, however, is the moment bound of Theorem 2.1 which turns out to be highly accurate: it has the optimal constant in its leading term, a fact that can be of significant importance in certain applications (an example of such an application will be given in section 4.3 below). If we wish to obtain an analogously sharp result in the non-symmetric case, the proof must be modified to account for the non-symmetric structure. This results in the following bound.

**Theorem 4.2** *Let $X$ be an $n\times m$ matrix with $X_{ij} = b_{ij}g_{ij}$, where $b_{ij}\ge 0$ and $g_{ij}$ are i.i.d. standard Gaussian variables for $i\le n$, $j\le m$. Then for $p\in\mathbb{N}$*

$$(\mathbf{E}\|X\|_{S_{2p}}^{2p})^{\frac{1}{2p}} \le \Bigg(\sum_i\Bigg(\sum_j b_{ij}^2\Bigg)^p\Bigg)^{\frac{1}{2p}} + \Bigg(\sum_j\Bigg(\sum_i b_{ij}^2\Bigg)^p\Bigg)^{\frac{1}{2p}} + 4\sqrt{p}\Bigg(\sum_{i,j} b_{ij}^{2p}\Bigg)^{\frac{1}{2p}}.$$

The adaptations of the proof of Theorem 2.1 needed to obtain this bound are analogous to the ones developed in [2, section 3.1] in the simpler setting considered there. The present case requires essentially no new ideas, but a complete rewriting of the combinatorial arguments of section 2 in the non-symmetric case would be very tedious. Instead, we will briefly sketch the necessary modifications in the rest of this section, leaving a detailed verification of the proofs to the interested reader.

*Proof (Sketch of proof of Theorem 4.2)* Define

$$\sigma_{p,1} := \Bigg(\sum_i\Bigg(\sum_j b_{ij}^2\Bigg)^p\Bigg)^{1/2p}, \qquad \sigma_{p,2} := \Bigg(\sum_j\Bigg(\sum_i b_{ij}^2\Bigg)^p\Bigg)^{1/2p}.$$

Throughout the proof, we may assume without loss of generality that $\sum_{i,j} b_{ij}^{2p} = 1$ (by scaling) and that $\sigma_{p,1}\ge\sigma_{p,2}$ (if not, replace $X$ by $X^*$). We begin by writing

$$\begin{aligned}\mathbf{E}\|X\|_{S_{2p}}^{2p} &= \mathbf{E}[\mathrm{Tr}[(XX^*)^p]] \\ &= \sum_{\mathbf{u}\in[n]^p}\sum_{\mathbf{v}\in[m]^p}\mathbf{E}[X_{u_1v_1}X_{u_2v_1}X_{u_2v_2}X_{u_3v_2}\cdots X_{u_pv_p}X_{u_1v_p}].\end{aligned}$$

We view $u_1 \to v_1 \to u_2 \to v_2 \to \cdots \to u_p \to v_p \to u_1$ as a cycle of length $2p$ in the complete undirected bipartite graph with $n$ left vertices and $m$ right vertices; we will refer to such a bipartite cycle for simplicity as a *bicycle*. By symmetry of the Gaussian distribution, it is clear that each distinct edge $\{u_k, v_k\}$ or $\{u_{k+1}, v_k\}$ that appears in the bicycle must be traversed an even number of times for a term in the sum to be nonzero; we will call bicycles with this property *even*.

The *shape* $\mathbf{s}(\mathbf{u}, \mathbf{v})$ of a bicycle $(\mathbf{u}, \mathbf{v}) \in [n]^p \times [m]^p$ is defined by relabeling the left and right vertices in order of appearance. Let $\mathcal{S}_{2p}^{\mathrm{bi}}$ be the set of shapes of even bicycles of length $2p$, denote by $n_i(\mathbf{s})$ the number of distinct edges that are traversed exactly $i$ times by $\mathbf{s}$, and denote by $m_1(\mathbf{s})$ and $m_2(\mathbf{s})$ the number of distinct right and left vertices, respectively, that are visited by $\mathbf{s}$. Then we can write

$$\mathbf{E}\|X\|_{S_{2p}}^{2p} = \sum_{\mathbf{s}\in\mathcal{S}_{2p}^{\mathrm{bi}}} \prod_{i\ge 1} \mathbf{E}[g^{2i}]^{n_{2i}(\mathbf{s})} \sum_{(\mathbf{u},\mathbf{v}):\mathbf{s}(\mathbf{u},\mathbf{v})=\mathbf{s}} b_{u_1v_1} b_{u_2v_1} \cdots b_{u_pv_p} b_{u_1v_p},$$

where $g \sim N(0, 1)$. The following is a non-symmetric analogue of Proposition 2.5.

**Proposition 4.3** *Suppose $\sum_{i,j} b_{ij}^{2p} = 1$ and $\sigma_{p,1} \ge \sigma_{p,2}$. Then for any $\mathbf{s} \in \mathcal{S}_{2p}^{\mathrm{bi}}$*

$$\sum_{(\mathbf{u},\mathbf{v}):\mathbf{s}(\mathbf{u},\mathbf{v})=\mathbf{s}} b_{u_1v_1} b_{u_2v_1} \cdots b_{u_pv_p} b_{u_1v_p} \le \sigma_{p,1}^{2m_1(\mathbf{s})} \sigma_{p,2}^{2(m_2(\mathbf{s})-1)}.$$

With this result in hand, the rest of the proof of Theorem 4.2 is almost identical to that of Theorem 2.1. Indeed, repeating the same steps, we can show that $\mathbf{E}[\mathrm{Tr}[(XX^*)^p]] \le \mathbf{E}[\|Y\|^{2p}]$, where $Y$ is the $r \times r'$ matrix whose entries are i.i.d. standard Gaussian with $r = \lceil \sigma_{p,2}^2 + p/2 \rceil$ and $r' = \lceil \sigma_{p,1}^2 + p/2 \rceil$. The conclusion of Theorem 4.2 now follows from a classical estimate on the norm of Gaussian matrices

$$\mathbf{E}[\|Y\|^{2p}]^{1/2p} \le \sqrt{r} + \sqrt{r'} + 2\sqrt{p}.$$

The omitted steps are spelled out in more detail in the proof of [2, Theorem 3.1]. □

What remains is to prove Proposition 4.3, which proceeds mostly along the lines of Proposition 2.5 but with additional bookkeeping issues. We will again merely sketch the necessary modifications to the proof of Proposition 2.5.

*Proof (Sketch of proof of Proposition 4.3)* Let $\mathcal{G}_{m_1,m_2}$ be the set of undirected, connected bipartite graphs with $m_1$ right vertices and $m_2$ left vertices. Exactly as in section 2.2, we associate to every shape $\mathbf{s} \in \mathcal{S}_{2p}^{\mathrm{bi}}$ with $m_1(\mathbf{s}) = m_1$ and $m_2(\mathbf{s}) = m_2$ a graph $G \in \mathcal{G}_{m_1,m_2}$ and weights $\mathbf{k} = (k_e)_{e\in E(G)}$, $k_e \ge 2$ with $|\mathbf{k}| = 2p$ such that

$$\sum_{(\mathbf{u},\mathbf{v}):\mathbf{s}(\mathbf{u},\mathbf{v})=\mathbf{s}} b_{u_1v_1} b_{u_2v_1} \cdots b_{u_pv_p} b_{u_1v_p} \le W^{\mathbf{k}}(G).$$

Here we define in the bipartite case

$$W^{\mathbf{k}}(G) := \sum_{\mathbf{u}\in[n]^{m_2}} \sum_{\mathbf{v}\in[m]^{m_1}} \prod_{e\in E(G)} b_{u(e)v(e)}^{k_e},$$

where we denote by $e = (u(e), v(e))$ the left and right vertices of the edge $e$. We must show that for every $G \in \mathcal{G}_{m_1,m_2}$ and $\mathbf{k} = (k_e)_{e\in E(G)}$, $k_e \geq 2$, $|\mathbf{k}| = 2p$ we have

$$W^{\mathbf{k}}(G) \leq \sigma_{p,1}^{2m_1}\sigma_{p,2}^{2(m_2-1)}$$

under the assumptions of Proposition 4.3.

We begin by observing that it suffices to prove the claim only for $G \in \mathcal{G}^{\mathrm{tree}}_{m_1,m_2}$, that is, when $G$ is a bipartite tree with $m_1$ right vertices and $m_2$ left vertices. The proof is identical to that of Lemma 2.9, so we do not comment on it further.

To obtain the requisite bound for trees, we now proceed exactly as in the proof of Lemma 2.10 by iteratively pruning the leaves of the tree using Hölder's inequality. However, in the present case the coefficient matrix $b_{ij}$ is not symmetric, so we must keep track of which of its indices is being summed over in each step of the iteration. As the same edge may be pruned from either direction in the course of the induction, the final conclusion of the induction argument is a bound of the form

$$W^{\mathbf{k}}(G) \leq \prod_{e\in E(G)} \Bigg[\sum_i \Bigg(\sum_j b_{ij}^{k_e}\Bigg)^{\frac{|\mathbf{k}|}{k_e}}\Bigg]^{\frac{1}{\alpha_e}} \Bigg[\sum_j \Bigg(\sum_i b_{ij}^{k_e}\Bigg)^{\frac{|\mathbf{k}|}{k_e}}\Bigg]^{\frac{1}{\beta_e}}.$$

However, recall that in each step of the induction argument, the homogeneity in each variable $b_{ij}^{k_e}$ is preserved. Moreover, in the non-symmetric case, it is readily seen that the total number of sums over left and right vertices, respectively, is also preserved in each step of the induction argument (for example, if we bound $\sum_{i,j,k} b_{ij}^2 b_{kj}^2 \leq [\sum_j(\sum_i b_{ij}^2)^2]^{1/2}[\sum_j(\sum_k b_{kj}^2)^2]^{1/2}$, there are two sums over left vertices and one sum over a right vertex on both sides of the inequality when we count multiplicities given by the exponents.) Thus the following must be valid for the final bound:

- The right-hand side is 1-homogeneous in each variable $b_{ij}^{k_e}$. Therefore, we must have $1/\alpha_e + 1/\beta_e = k_e/|\mathbf{k}|$ for every $e \in E(G)$.
- There are $m_1$ sums over right vertices and $m_2$ sums over left vertices on the right-hand side (counting multiplicities given by the exponents). Therefore, we must have $\sum_{e\in E(G)}(|\mathbf{k}|/k_e\alpha_e + 1/\beta_e) = m_1$ and $\sum_{e\in E(G)}(|\mathbf{k}|/k_e\beta_e + 1/\alpha_e) = m_2$.

Now apply Hölder's inequality to each term as we did at the end of the proof of Theorem 2.8. This yields, using $\sum_{i,j} b_{ij}^{|\mathbf{k}|} = \sum_{i,j} b_{ij}^{2p} = 1$,

$$W^{\mathbf{k}}(G) \leq \prod_{e\in E(G)} \Bigg[\sum_i \Bigg(\sum_j b_{ij}^{2}\Bigg)^{\frac{|\mathbf{k}|}{2}}\Bigg]^{\frac{2}{k_e\alpha_e}} \Bigg[\sum_j \Bigg(\sum_i b_{ij}^{2}\Bigg)^{\frac{|\mathbf{k}|}{2}}\Bigg]^{\frac{2}{k_e\beta_e}}.$$

But as $\sigma_{p,1} \geq \sigma_{p,2}$, we can estimate

$$W^{\mathbf{k}}(G) = \sigma_{p,1}^{2\sum_e \frac{|\mathbf{k}|}{k_e\alpha_e}}\sigma_{p,2}^{2\sum_e \frac{|\mathbf{k}|}{k_e\beta_e}} \leq \sigma_{p,1}^{2\sum_e(\frac{|\mathbf{k}|}{k_e\alpha_e}+\frac{1}{\beta_e})}\sigma_{p,2}^{2\sum_e(\frac{|\mathbf{k}|}{k_e\beta_e}-\frac{1}{\beta_e})} = \sigma_{p,1}^{2m_1}\sigma_{p,2}^{2(m_2-1)},$$

which completes the proof. $\square$

### 4.2 Non-Gaussian matrices: heavy-tailed entries

The main results of this paper show that for a centered Gaussian random matrix $X$, the distributions of the random variables $\|X\|_{S_p}$ and $\|X\|_{\ell_p(\ell_2)}$ are comparable in a very strong sense. This conclusion is surprisingly general—the comparison holds regardless of the variance pattern of the matrix entries—and it is far from clear from its simple statement why even Gaussianity of the entries should be needed. It is tempting to conjecture that this phenomenon should not depend on the entry distribution at all, but should follow from a general comparison principle for arbitrary random matrices with independent entries. However, this is not the case: as is shown by a simple example due to Seginer [15, Theorem 3.2], even the conclusion $\mathbf{E}\|X\|_{S_p} \lesssim \mathbf{E}\|X\|_{\ell_p(\ell_2)}$ is manifestly false when the entries of $X$ are uniformly bounded. From this perspective, it is not a surprise that the sharp result proves to be a Gaussian phenomenon, and that Gaussian analysis should play a key role throughout its proof.

Nonetheless, our results are by no means restricted to Gaussian entries: once the Gaussian result has been proved, we can extend its conclusion to a much larger class of distributions. To fix some ideas, consider for example the case where $X$ is a symmetric matrix with $X_{ij} = b_{ij}h_{ij}$, where $b_{ij} \geq 0$ and $h_{ij}$ are i.i.d. symmetric $\alpha$-stable random variables with $\alpha \in (1,2]$ (for $i \geq j$). We claim that the conclusion of Corollary 1.4 extends verbatim to this situation: that is, we have

$$\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \geq t] \leq \mathbf{P}[\|X\|_{S_p} \geq t] \leq C\,\mathbf{P}[\|X\|_{\ell_p(\ell_2)} \geq t/C]$$

for all $t \geq 0$ and $2 \leq p \leq \infty$, where $C$ is a universal constant. To prove this, it suffices to recall [6, p. 176] that any symmetric $\alpha$-stable random variable can be written as $h_{ij} = g_{ij}z_{ij}$, where $g_{ij}$ is a standard Gaussian variable and $z_{ij}$ is a nonnegative random variable independent of $g_{ij}$. It therefore suffices to apply Corollary 1.4 conditionally on $\{z_{ij}\}$ to prove the claim.

The above observations suggest the following general principle: while the conclusion of Theorem 3.1 (or Corollary 1.4) cannot be expected to hold in general for light-tailed entries, the result should extend rather generally to entry distributions whose tails are heavier than that of the Gaussian distribution. We have not proved a completely general formulation of this idea, and it is unclear what minimal regularity requirements are needed to make it precise. However, the following partial result applies to a broad class of heavy-tailed entry distributions, and serves as an illustration of how our results may be extended far beyond the Gaussian setting.

**Theorem 4.4** *Let $X$ be an $n \times n$ symmetric matrix with $X_{ij} = b_{ij}h_{ij}$, where $b_{ij} \geq 0$ and $h_{ij}$, $i \geq j$ are independent centered random variables that satisfy*

$$C_1 p^\beta \leq \mathbf{E}[|h_{ij}|^p]^{1/p} \leq C_2 p^\beta \qquad \text{for all } p \geq 2$$

*for some $\beta \geq \frac{1}{2}$. Then for $2 \leq p \leq \infty$*

$$\begin{aligned}\mathbf{E}\|X\|_{S_p} &\asymp \mathbf{E}\Bigg[\bigg(\sum_i \bigg(\sum_j X_{ij}^2\bigg)^{p/2}\bigg)^{1/p}\Bigg] \\ &\asymp \bigg(\sum_i \bigg(\sum_j b_{ij}^2\bigg)^{p/2}\bigg)^{1/p} + \max_{i\leq e^p}\max_j b_{ij}^*(\log i)^\beta + p^\beta\bigg(\sum_{i\geq e^p}\max_j {b_{ij}^*}^p\bigg)^{1/p},\end{aligned}$$

*where the constants depend on $C_1, C_2, \beta$ only.*

**Remark 4.5** One may readily verify by inspection of the proof that the conclusions of Corollaries 1.4 and 4.1 extend analogously to the setting of Theorem 4.4.

The rest of this subsection is devoted to the proof of Theorem 4.4. As a first step, we note that the moment assumption implies a tail bound.

**Lemma 4.6** *Let $h$ be a random variable such that*

$$C_1 p^\beta \le \mathbf{E}[|h|^p]^{1/p} \le C_2 p^\beta \qquad \text{for all } p \ge 2.$$

*Then there exist constants $c_1, c_2$ depending only on $C_1, C_2, \beta$ such that*

$$c_1 e^{-t^{1/\beta}/c_1} \le \mathbf{P}[|h| \ge t] \le c_2 e^{-t^{1/\beta}/c_2} \qquad \text{for all } t \ge 0.$$

*Proof* To upper bound the tail, we note that Markov's inequality implies

$$\mathbf{P}[|h| \ge C_2 e p^\beta] \le \mathbf{P}[|h| \ge e\|h\|_p] \le e^{-p}$$

for all $p \ge 2$. Moreover, if we further bound $e^{-p}$ by $e^{2-p}$, then the inequality is trivially valid for all $p > 0$. The upper tail now follows easily.

To lower bound the tail, we note that by the Paley-Zygmund inequality

$$\mathbf{P}[|h| \ge \tfrac{1}{2} C_1 p^\beta] \ge \mathbf{P}[|h| \ge \tfrac{1}{2}\|h\|_p] \ge (1 - 2^{-p})^2 \left( \frac{\|h\|_p}{\|h\|_{2p}} \right)^{2p} \ge \tfrac{1}{2} e^{-cp}$$

for $p \ge 2$, where $c = 2\log(2^\beta C_2/C_1) > 0$. Moreover, if we lower bound $e^{-cp}$ by $e^{-c(p+2)}$, the inequality is valid for all $p > 0$. The lower tail follows easily. □

As a consequence, we obtain the following comparison theorem.

**Lemma 4.7** *Let $h_i$ and $h_i'$, $i = 1, \ldots, k$ be independent centered random variables that satisfy the moment assumption of Lemma 4.6 with constants $C_1, C_2, \beta$. Then there exists a constant $C$ depending only on $C_1, C_2, \beta$ such that*

$$\mathbf{E}[f(h_1, \ldots, h_k)] \le \mathbf{E}[f(Ch_1', \ldots, Ch_k')]$$

*for every symmetric convex function $f : \mathbb{R}^k \to \mathbb{R}$.*

*Proof* We begin by noting that as $h_i$ are centered, Jensen's inequality yields

$$\mathbf{E}[f(h_1, \ldots, h_k)] \le \mathbf{E}[f(h_1 - \tilde{h}_1, \ldots, h_k - \tilde{h}_k)] \le \mathbf{E}[f(2h_1, \ldots, 2h_k)]$$

whenever $f$ is symmetric and convex, where $\tilde{h}_i$ are independent copies of $h_i$. Moreover, the random variables $h_i - \tilde{h}_i$ clearly satisfy the same moment assumptions as $h_i$ modulo a universal constant. We can therefore assume without loss of generality in the sequel that the random variables $h_i$ and $h_i'$ are symmetrically distributed.

To proceed, note that Lemma 4.6 implies that

$$\mathbf{P}[|h_i| \ge t] \le c\, \mathbf{P}[c|h_i'| \ge t]$$

for all $t \geq 0$ and $i$, where $c \geq 1$ is a constant that depends only on $C_1, C_2, \beta$. In particular, let $\delta_i \sim \mathrm{Bern}(1/c)$ be i.i.d. Bernoulli variables independent of $h$. Then

$$\mathbf{P}[\delta_i|h_i| \geq t] \leq \mathbf{P}[c|h'_i| \geq t]$$

for all $t \geq 0$. By a standard coupling argument, we can couple $(\delta_i, h_i)$ and $(h'_i)$ on the same probability space such that $\delta_i|h_i| \leq c|h'_i|$ a.s. for every $i$ [10, p. 127].

Now let $\varepsilon_i$ be i.i.d. Rademacher variables. Then

$$\mathbf{E}[f(\varepsilon_1\delta_1|h_1|, \ldots, \varepsilon_k\delta_k|h_k|)\,|\delta, h, h'] \leq \mathbf{E}[f(c\varepsilon_1|h'_1|, \ldots, c\varepsilon_k|h'_k|)\,|\delta, h, h']$$

follows by convexity (it suffices to note that $\alpha \mapsto \mathbf{E}[f(\alpha_1\varepsilon_1, \ldots, \alpha_k\varepsilon_k)]$ is convex, so its supremum over $\prod_i[-c|h'_i|, c|h'_i|]$ is attained at one of the extreme points). Therefore, as $h_i, h'_i$ are symetrically distributed and by Jensen's inequality,

$$\mathbf{E}[f(h_1/c, \ldots, h_k/c)] \leq \mathbf{E}[f(\delta_1 h_1, \ldots, \delta_k h_k)] \leq \mathbf{E}[f(ch'_1, \ldots, ch'_k)]$$

for every symmetric convex function $f$. This concludes the proof. □

We can now complete the proof of Theorem 4.4.

*Proof (Proof of Theorem 4.4)* Let $g_{ij}$ and $\tilde{g}_{ij}$, $i \geq j$ be i.i.d. standard Gaussian variables, and define $h'_{ij} := g_{ij}|\tilde{g}_{ij}|^{2\beta-1}$. Then it is readily verified that $h'_{ij}$ satisfies the same moment condition as $h_{ij}$ for each $i, j$. Therefore, by Lemma 4.7, we have

$$\mathbf{E}\|X\|_{S_p} \asymp \mathbf{E}\|(b_{ij}h'_{ij})\|_{S_p}, \qquad \mathbf{E}\|X\|_{\ell_p(\ell_2)} \asymp \mathbf{E}\|(b_{ij}h'_{ij})\|_{\ell_p(\ell_2)},$$

where the constants depend only on $C_1, C_2, \beta$. By applying Theorem 3.1 conditionally on $\{\tilde{g}_{ij}\}$, it now follows immediately that $\mathbf{E}\|X\|_{S_p} \asymp \mathbf{E}\|X\|_{\ell_p(\ell_2)}$.

It remains to obtain the explicit formula. To this end, let $\varepsilon_{ij}$, $i \geq j$ be i.i.d. Rademacher variables and let $h''_{ij} := \varepsilon_{ij}|g_{ij}|^{2\beta}$. Applying again Lemma 4.7 yields

$$\mathbf{E}\|X\|_{\ell_p(\ell_2)} \asymp \mathbf{E}\|(b_{ij}h''_{ij})\|_{\ell_p(\ell_2)} = \mathbf{E}\|(b_{ij}^{1/2\beta}g_{ij})\|^{2\beta}_{\ell_{2\beta p}(\ell_{4\beta})}.$$

A routine application of Gaussian concentration [5, Theorem 5.8] shows that

$$\mathbf{E}[\|(b_{ij}^{1/2\beta}g_{ij})\|^{2\beta}_{\ell_{2\beta p}(\ell_{4\beta})}]^{1/2\beta} \asymp \mathbf{E}\|(b_{ij}^{1/2\beta}g_{ij})\|_{\ell_{2\beta p}(\ell_{4\beta})},$$

where the constant depends only on $\beta$. The proof is concluded by a straightforward adaptation of the proof of Corollary 3.6; we omit the details. □

### 4.3 Non-Gaussian matrices: bounded entries

As was explained in the previous section, the two-sided bounds of Theorem 3.1 fail to extend to the situation where the entries of the matrix are light-tailed; in such cases new phenomena arise that are poorly understood, and the problem of obtaining two-sided bounds for matrices with bounded entries remains open (see [2, section 4.2] for some discussion along these lines). Nonetheless, Gaussian results are still of considerable interest in this setting as they yield very good upper bounds on the matrix norms in many cases of practical interest. For example, the methods of the previous subsection may be easily adapted to show that if $X$ is a symmetric random matrix whose entries $X_{ij}$ are independent, centered, and $b_{ij}$-subgaussian, then its expected Schatten norms are dominated by those of the Gaussian matrix defined in Theorem 3.1. Thus the explicit expression given in Theorem 3.1, while not always sharp in the subgaussian case, always yields an upper bound on the quantities of interest.

Of particular interest in this context are the bounds of Theorems 2.1 and 4.2, which not only yield very explicit upper bounds on the moments of Gaussian random matrices, but even provide sharp constants in the leading terms. The aim of this section is to show that for random matrices with uniformly bounded entries, these moment bounds admit an important refinement. While this requires only a minor modification of the proofs of Theorems 2.1 and 4.2, we will spell out these results in some detail as they prove to be of considerable utility in many applications (for example, in the study of random graphs and in applied mathematics).

The main result of this section is the following slight refinement of Theorem 1.5.

**Theorem 4.8** *Let $X$ be an $n \times n$ symmetric matrix with independent centered entries for $i \geq j$, and define the quantities*

$$\sigma_p := \left( \sum_i \left( \sum_j \mathbf{E}[X_{ij}^2] \right)^p \right)^{1/2p}, \qquad \sigma_p^* := \left( \sum_{i,j} \|X_{ij}\|_\infty^{2p} \right)^{1/2p}.$$

*Then we have for every $p \in \mathbb{N}$*

$$(\mathbf{E}\|X\|_{S_{2p}}^{2p})^{1/2p} \leq 2\sigma_p + C\sqrt{p}\,\sigma_p^*,$$

*where $C$ is a universal constant. Moreover, we have*

$$\mathbf{P}[\|X\|_{S_{2p}} \geq 2\sigma_p + C\sqrt{p}\,\sigma_p^* + t] \leq e^{-t^2/C\max_{ij}\|X_{ij}\|_\infty^2}$$

*for all $p \in \mathbb{N}$ and $t \geq 0$.*

A non-symmetric analogue will follow in precisely the same manner.

**Theorem 4.9** *Let $X$ be an $n \times m$ matrix with independent centered entries. Define*

$$\sigma_{p,1} := \left( \sum_i \left( \sum_j \mathbf{E}[X_{ij}^2] \right)^p \right)^{1/2p}, \qquad \sigma_{p,2} := \left( \sum_j \left( \sum_i \mathbf{E}[X_{ij}^2] \right)^p \right)^{1/2p},$$

*and let $\sigma_p^*$ be defined as in Theorem 4.8. Then we have for every $p \in \mathbb{N}$*

$$(\mathbf{E}\|X\|_{S_{2p}}^{2p})^{1/2p} \le \sigma_{p,1} + \sigma_{p,2} + C\sqrt{p}\,\sigma_p^*,$$

*where C is a universal constant. Moreover, we have*

$$\mathbf{P}[\|X\|_{S_{2p}} \ge \sigma_{p,1} + \sigma_{p,2} + C\sqrt{p}\,\sigma_p^* + t] \le e^{-t^2/C\max_{ij}\|X_{ij}\|_\infty^2}$$

*for all $p \in \mathbb{N}$ and $t \ge 0$.*

The key point in these results is that the leading term only depends on the variances of the matrix entries, while the second term depends on their uniform bound. The variance-sensitive nature of the bound is crucial in many applications. For example, we briefly describe an application to random graphs.

*Example 4.1* Let $A$ be the adjacency matrix of a nonhomogeneous Erdős-Rényi random graph on $n$ vertices, where each edge $\{i, j\}$ is included independently with probability $p_{ij}$. Applying Theorem 4.8 with $p = \lceil \alpha \log n \rceil$ yields

$$\mathbf{E}\|A - \mathbf{E}A\| \le (\mathbf{E}\|A - \mathbf{E}A\|_{S_{2p}}^{2p})^{1/2p} \le 2e^{1/2\alpha}\sqrt{d} + Ce^{1/\alpha}\sqrt{\alpha \log n}$$

for any $\alpha \ge 1$, where we defined $d := \max_i \sum_j p_{ij}$. It follows immediately that

$$\frac{\mathbf{E}\|A - \mathbf{E}A\|}{\sqrt{d}} \le 2(1 + o(1)) \qquad \text{when } n \to \infty,\ \log n = o(d).$$

We therefore easily recover a recent result of [3] that was obtained there by a more complicated method. For the purpose of this application, it is important to note that both the constant 2 and the condition $d \gtrsim \log n$ are in fact optimal, at least in the homogeneous case; see [3] and the references therein. Our general bounds therefore yield surprisingly accurate results in this example.

**Remark 4.10** If the entries $X_{ij}$ are symmetrically distributed, one can deduce the results of Theorems 4.8 and 4.9 directly from Theorems 2.1 and 4.2 by a simple symmetrization argument, cf. [2, Corollary 3.6]. However, if we only assume that $X_{ij}$ are centered, the symmetrization method loses an additional factor $\sqrt{2}$, while the sharp constant for non-symmetrically distributed entries was essential in the application of Example 4.1. The main observation of this section, which is implicitly contained but not stated in [2], is that a minor modification of the proof of the moment bounds makes it possible to obtain the optimal constant even when the entries are only assumed to be centered. For the purpose of Example 4.1, it would suffice to apply this idea in the simpler setting developed in [2].

**Remark 4.11** As was already used implicitly in Example 4.1, Theorems 4.8 and 4.9 provide rather practical bounds on the operator norm of $X$ by choosing $p \sim \log n$. For example, applying Theorem 4.8 with $p = \lceil \alpha \log n \rceil$ and using $e^{1/\alpha} \le 1 + 2/\alpha$ for $\alpha \ge 1$, we can deduce as in the proof of [2, Corollary 3.12] that

$$\mathbf{P}\left[\|X\| \ge 2(1+\varepsilon)\max_i \sqrt{\sum_j \mathbf{E}[X_{ij}^2]} + t\right] \le ne^{-\varepsilon t^2/C\max_{ij}\|X_{ij}\|_\infty^2}$$

for every $t \geq 0$ and $0 \leq \varepsilon \leq 1$ ($C$ is a universal constant). This result improves on [2, Corollary 3.12] in that it attains the optimal constant 2 in the probability bound assuming only that the entries $X_{ij}$ are centered, rather than symmetrically distributed. For non-symmetric matrices, we obtain analogously that

$$\mathbf{P}\left[\|X\| \geq (1+\varepsilon)\left(\max_i \sqrt{\sum_j \mathbf{E}[X_{ij}^2]} + \max_j \sqrt{\sum_i \mathbf{E}[X_{ij}^2]}\right) + t\right] \leq \max(n,m) e^{-\varepsilon t^2/C \max_{ij} \|X_{ij}\|_\infty^2}$$

in the setting of Theorem 4.9. Such "matrix concentration inequalities" have found numerous applications in applied mathematics (see [17] and the references therein).

The rest of this section is devoted to the proof of Theorem 4.8. We omit the proof of Theorem 4.9, which follows in an identical manner.

To explain the idea behind the proof, let us make a basic observation: while Gaussian analysis played a crucial role for the norm bound of Theorem 3.1, the entry distribution was completely irrelevant in the moment bound of Theorem 2.1. Indeed, all the proof does is to compare the moments of the nonhomogeneous random matrix $X$ with the moments of another random matrix $Y$ that has i.i.d. entries. If $X$ is Gaussian, then we may choose $Y$ to be Gaussian as well, and we conclude by invoking standard bounds on the norm of Gaussian Wigner matrices. If the entries of $X$ are bounded and centered, exactly the same argument will apply provided we select an appropriate entry distribution for the matrix $Y$. Modulo this minor observation, the rest of the proof transfers readily to the present setting.

*Proof (Proof of Theorem 4.8)* As in the proof of Theorem 2.1, we will assume without loss of generality that the diagonal entries of the matrix vanish $X_{ii} = 0$.

Our starting point is again the moment formula

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] = \sum_{\mathbf{u}\in[n]^{2p}} \mathbf{E}[X_{u_1u_2}X_{u_2u_3}\cdots X_{u_{2p}u_1}].$$

Each distinct edge $\{u_k, u_{k_1}\}$ that appears in the cycle $\mathbf{u}$ must be traversed at least twice for that term in the sum to be nonzero, as we assumed the random variables $X_{ij}$ are centered. We call a cycle with this property *admissible*. Note that unlike in the Gaussian setting of section 2, an admissible cycle is not necessarily even. This distinction will turn out to require only minimal modifications to the proof.

Let us denote by $\tilde{\mathcal{S}}_{2p}$ the set of shapes of admissible cycles of length $2p$. Then

$$\begin{aligned}\mathbf{E}[\mathrm{Tr}[X^{2p}]] &= \sum_{\mathbf{s}\in\tilde{\mathcal{S}}_{2p}} \sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} \mathbf{E}[X_{u_1u_2}X_{u_2u_3}\cdots X_{u_{2p}u_1}] \\ &\leq \sum_{\mathbf{s}\in\tilde{\mathcal{S}}_{2p}} \sum_{\mathbf{v}\in[n]^{m(\mathbf{s})}} \prod_{e\in E(G(\mathbf{s}))} \mathbf{E}[|X_{v(e)}|^{k_e(\mathbf{s})}],\end{aligned}$$

where $m(\mathbf{s})$ is the number of distinct vertices visited by $\mathbf{s}$, $G(\mathbf{s}) \in \mathcal{G}_{m(\mathbf{s})}$ is the graph whose edges are given by $E(G(\mathbf{s})) = \{\{s_1, s_2\}, \{s_2, s_3\}, \ldots, \{s_{2p}, s_1\}\}$, and $k_e(\mathbf{s})$ is the number of times edge $e \in E(G(\mathbf{s}))$ is traversed by $\mathbf{s}$.

We can now easily adapt the proof of Theorem 2.8 to show that for every $\mathbf{s} \in \tilde{\mathcal{S}}_{2p}$, there exist $k'_1, \ldots, k'_m \geq 2$ with $\sum_{\ell=2}^m k'_\ell = 2p$ such that

$$\sum_{\mathbf{v}\in[n]^{m(\mathbf{s})}} \prod_{e\in E(G(\mathbf{s}))} \mathbf{E}[|X_{v(e)}|^{k_e(\mathbf{s})}] \leq \prod_{\ell=2}^{m(\mathbf{s})} \left( \sum_i \left( \sum_j \mathbf{E}[|X_{ij}|^{k'_\ell}] \right)^{\frac{2p}{k'_\ell}} \right)^{\frac{k'_\ell}{2p}}.$$

Indeed, if we simply replace $b_{ij}^k$ by $\mathbf{E}[|X_{ij}|^k]$ throughout, Lemma 2.9 extends directly to the present setting (modulo a trivial application of Jensen's inequality $\mathbf{E}[|X_{ij}|^{k(i)}] \leq \mathbf{E}[|X_{ij}|^{k'_m}]^{k(i)/k'_m}$ in the second and in the last equation display of the proof), while Lemma 2.10 can be applied verbatim. We further estimate

$$\sum_i \left( \sum_j \mathbf{E}[|X_{ij}|^{k'_\ell}] \right)^{\frac{2p}{k'_\ell}} \leq \sum_i \left( \sum_j \mathbf{E}[|X_{ij}|^2] \right)^{\frac{2p}{k'_\ell}} \max_j \|X_{ij}\|_\infty^{\frac{2p(k'_\ell-2)}{k'_\ell}} \leq \sigma_p^{\frac{4p}{k'_\ell}} (\sigma_p^*)^{2p-\frac{4p}{k'_\ell}}$$

using Hölder's inequality, where $\sigma_p, \sigma_p^*$ are as defined in Theorem 4.8.

By rescaling the matrix $X$, we may assume without loss of generality that $\sigma_p^* = 1$. Putting together the above estimates, we have shown in this case that

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] \leq \sum_{\mathbf{s}\in\tilde{\mathcal{S}}_{2p}} \sigma_p^{2(m(\mathbf{s})-1)}.$$

Now let $Y$ be an $r \times r$ symmetric matrix with entries $Y_{ij} = (\delta_{ij} - \mathbf{E}[\delta_{ij}])/\mathrm{Var}(\delta_{ij})^{1/2}$, where $\delta_{ij} \sim \mathrm{Bern}(\frac{1}{4})$ are i.i.d. Bernoulli random variables for $i \geq j$. The key point of this choice of distribution is that $\mathbf{E}[Y_{ij}^k] \geq 1$ for every integer $k \geq 2$, as may be verified by a simple explicit computation. We therefore obtain

$$\mathbf{E}[\mathrm{Tr}[Y^{2p}]] = \sum_{\mathbf{s}\in\tilde{\mathcal{S}}_{2p}} \sum_{\mathbf{u}:\mathbf{s}(\mathbf{u})=\mathbf{s}} \mathbf{E}[Y_{u_1u_2}Y_{u_2u_3}\cdots Y_{u_{2p}u_1}] \geq \sum_{\mathbf{s}\in\tilde{\mathcal{S}}_{2p}} \frac{r!}{(r-m(\mathbf{s}))!}.$$

It now follows exactly as in the proof of Theorem 2.1 that

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]] \leq \mathbf{E}[\|Y\|^{2p}] \quad \text{for} \quad r = \lfloor \sigma_p^2 \rfloor + p + 1.$$

To control $\|Y\|$, it remains to invoke a norm bound for Wigner matrices with non-symmetrically distributed entries. Such a bound can be found, for example, in [12], where it is shown that $\mathbf{E}\|Y\| \leq 2\sqrt{r} + C'$ for a universal constant $C'$. Applying Talagrand's concentration inequality [5, Theorem 6.10], we can conclude that

$$\mathbf{E}[\|Y\|^{2p}]^{1/2p} \leq 2\sqrt{r} + C''\sqrt{p}$$

for a universal constant $C''$. Therefore

$$\mathbf{E}[\mathrm{Tr}[X^{2p}]]^{1/2p} \leq 2\sigma_p + C\sqrt{p},$$

concluding the proof of the moment bound. The tail bound follows immediately by another application of Talagrand's concentration inequality. □

**Remark 4.12** By a further adaptation of the method of proof of Theorem 4.8, one can prove general bounds that apply even to unbounded entries: for example, if $X$ is a symmetric matrix with independent centered entries for $i \geq j$, we have

$$(\mathbf{E}\|X\|_{S_{2p}}^{2p})^{1/2p} \leq 2\biggl(\mathbf{E}\biggl[\sum_i\biggl(\sum_j X_{ij}^2\biggr)^p\biggr]\biggr)^{1/2p} + C\sqrt{p}\biggl(\mathbf{E}\biggl[\sum_i \max_j X_{ij}^{2p}\biggr]\biggr)^{1/2p}$$

for all $p \in \mathbb{N}$. This could be viewed as a random matrix analogue of Rosenthal's inequality [5, Theorem 15.10]. Unfortunately, this inequality typically does not give the correct scaling in $p$ when the entries are unbounded (consider, for example, the case when the random variables are Gaussian). Nonetheless, such inequalities can be useful as they apply to very general situations without any additional work.

To prove this inequality, one simply estimates using Hölder's inequality

$$\sum_i\biggl(\sum_j \mathbf{E}[|X_{ij}|^{k'_\ell}]\biggr)^{\frac{2p}{k'_\ell}} \leq \biggl(\mathbf{E}\biggl[\sum_i\biggl(\sum_j X_{ij}^2\biggr)^p\biggr]\biggr)^{\frac{2}{k'_\ell}}\biggl(\mathbf{E}\biggl[\sum_i \max_j X_{ij}^{2p}\biggr]\biggr)^{1-\frac{2}{k'_\ell}}$$

instead of the corresponding estimate in the proof of Theorem 4.8.

**Acknowledgements** R.L. was supported by the National Science Centre of Poland under grant number 2015/18/A/ST1/00553. R.v.H. was supported in part by NSF grant CAREER-DMS-1148711 and by the ARO through PECASE award W911NF-14-1-0094. P.Y. was supported by grant ANR-16-CE40-0024-01. P.Y. would like to thank O. Guédon for introducing him to the questions investigated in this paper.

This work was conducted while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, supported by NSF grant DMS-1440140. The hospitality of MSRI and of the organizers of the program on Geometric Functional Analysis is gratefully acknowledged.

## References

1. Akhiezer, N.I., Glazman, I.M.: Theory of linear operators in Hilbert space. Dover Publications, Mineola, NY (1993)
2. Bandeira, A.S., van Handel, R.: Sharp nonasymptotic bounds on the norm of random matrices with independent entries. Ann. Probab. **44**(4), 2479–2506 (2016)
3. Benaych-Georges, F., Bordenave, C., Knowles, A.: Spectral radii of sparse random matrices (2017). Preprint arxiv:1704.02945
4. Bennett, J., Carbery, A., Christ, M., Tao, T.: Finite bounds for Hölder-Brascamp-Lieb multilinear inequalities. Math. Res. Lett. **17**(4), 647–666 (2010)
5. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities. Oxford University Press, Oxford (2013)
6. Feller, W.: An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons, Inc., New York-London-Sydney (1971)
7. van Handel, R.: On the spectral norm of Gaussian random matrices. Trans. Amer. Math. Soc. **369**(11), 8161–8178 (2017)
8. Landau, H.J., Shepp, L.A.: On the supremum of a Gaussian process. Sankhyā Ser. A **32**, 369–378 (1970)
9. Latała, R.: Some estimates of norms of random matrices. Proc. Amer. Math. Soc. **133**(5), 1273–1282 (electronic) (2005)
10. Lindvall, T.: Lectures on the coupling method. Dover Publications, Mineola, NY (2002)
11. Maligranda, L.: A property of interpolation spaces. Arch. Math. (Basel) **48**(1), 82–84 (1987)
12. Péché, S., Soshnikov, A.: Wigner random matrices with non-symmetrically distributed entries. J. Stat. Phys. **129**(5-6), 857–884 (2007)

13. Pisier, G.: Martingales in Banach spaces, *Cambridge Studies in Advanced Mathematics*, vol. 155. Cambridge University Press, Cambridge (2016)
14. Pisier, G., Xu, Q.: Non-commutative $L^p$-spaces. In: Handbook of the geometry of Banach spaces, Vol. 2, pp. 1459–1517. North-Holland, Amsterdam (2003)
15. Seginer, Y.: The expected norm of random matrices. Combin. Probab. Comput. **9**(2), 149–166 (2000)
16. Tao, T.: Topics in random matrix theory, *Graduate Studies in Mathematics*, vol. 132. American Mathematical Society, Providence, RI (2012)
17. Tropp, J.A.: An introduction to matrix concentration inequalities. Foundations and Trends in Machine Learning **8**, 1–230 (2015)