arXiv:2108.06312v1 [math.PR] 13 Aug 2021

MATRIX CONCENTRATION INEQUALITIES AND FREE PROBABILITY

AFONSO S. BANDEIRA, MARCH T. BOEDIHARDJO, AND RAMON VAN HANDEL

Abstract. A central tool in the study of nonhomogeneous random matrices, the noncommutative Khintchine inequality of Lust-Piquard and Pisier, yields a nonasymptotic bound on the spectral norm of general Gaussian random matrices $X = \sum_i g_i A_i$ where g_i are independent standard Gaussian variables and A_i are matrix coefficients. This bound exhibits a logarithmic dependence on dimension that is sharp when the matrices A_i commute, but often proves to be suboptimal in the presence of noncommutativity.

In this paper, we develop nonasymptotic bounds on the spectrum of arbitrary Gaussian random matrices that can capture noncommutativity. These bounds quantify the degree to which the deterministic matrices A_i behave as though they are freely independent. This "intrinsic freeness" phenomenon provides a powerful tool for the study of various questions that are outside the reach of classical methods of random matrix theory. Our nonasymptotic bounds are easily applicable in concrete situations, and yield sharp results in examples where the noncommutative Khintchine inequality is suboptimal. When combined with a linearization argument, our bounds imply strong asymptotic freeness (in the sense of Haagerup-Thorbjørnsen) for a remarkably general class of Gaussian random matrix models, including matrices that may be very sparse and that lack any special symmetries.

Beyond the Gaussian setting, we develop matrix concentration inequalities that capture noncommutativity for general sums of independent random matrices, which arise in many problems of pure and applied mathematics.

1. Introduction

The study of the spectrum of random matrices arises as a central problem in many areas of mathematics. Motivated by topics ranging from mathematical physics to operator algebras, much of classical random matrix theory is concerned with the study of highly homogeneous matrix ensembles, such as those with i.i.d. entries or that are invariant under symmetry groups. Deep results obtained over the past six decades by numerous mathematicians have resulted in a very detailed understanding of the asymptotic properties of such models [2, 33].

In contrast, many problems in areas such as functional analysis [12, 28] and in applied mathematics [35, 4] fall outside the scope of classical random matrix theory. The random matrix models that arise in such problems possess two common features. On the one hand, such models are often highly nonhomogeneous and lack any natural symmetries. On the other hand, the type of questions that arise in these areas are generally nonasymptotic in nature, as the study of nonhomogeneous models often does not lend itself naturally to an asymptotic formulation.

2010 *Mathematics Subject Classification.* 60B20; 60E15; 46L53; 46L54; 15B52.

Key words and phrases. Random matrices; matrix concentration inequalities; free probability.

The above considerations motivate the need for nonasymptotic methods that can capture the spectral properties of essentially arbitrarily structured nonhomogeneous random matrices. It may appear hopeless at first sight that anything at all can be said at this level of generality. Nonetheless, as we will recall below, there exists a set of tools, known colloquially as "matrix concentration inequalities", that makes it possible to compute certain spectral statistics of very general nonhomogeneous random matrices up to logarithmic factors in the dimension. The results of this paper provide a powerful refinement of this theory that makes it possible to achieve sharp results in many situations that are outside the reach of classical methods.

1.1. **Matrix concentration inequalities.** As a guiding motivation for this paper, consider the problem of estimating the spectral norm (i.e., largest singular value) of an arbitrary $d \times d$ self-adjoint random matrix with centered jointly Gaussian entries. Any such matrix X can be represented as

$$X = \sum_{i=1}^{n} g_i A_i, \tag{1.1}$$

where $A_i \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ are deterministic self-adjoint $d \times d$ matrices and g_i are i.i.d. standard real Gaussian variables. As was noted in [30], the noncommutative Khintchine inequality of Lust-Piquard and Pisier [28, §9.8] implies that[1]

$$\sigma(X) \lesssim \mathbf{E}\|X\| \lesssim \sigma(X)\sqrt{\log d}, \tag{1.2}$$

where we define

$$\sigma(X)^2 = \|\mathbf{E}X^2\| = \left\| \sum_{i=1}^{n} A_i^2 \right\|. \tag{1.3}$$

Thus the expected spectral norm of any Gaussian random matrix can be explicitly computed up to a logarithmic factor in the dimension.

It should be emphasized that (1.1) is an extremely general model: no assumption is made on the covariance of the entries of X, so that the model can capture arbitrary variance profiles and dependencies between the entries. Analogues of (1.2) extend even further to the model $X = \sum_i X_i$ where X_i are arbitrary independent random matrices. Due to their generality and ease of use, these "matrix concentration inequalities" [35] have had a major impact on numerous applications. On the other hand, the utility of (1.2) is limited by the gap between the upper and lower bounds, which becomes increasingly severe in high dimension.

To understand the origin of this gap, it is instructive to recall the basic principle behind the proofs of almost all known matrix concentration inequalities: the norm of a random matrix is largest when the coefficients A_i commute. This idea arises clearly in proofs of these inequalities [35, 36, 39]: the key step is application of trace inequalities that permute the order of the matrices A_i, which become equalities when all A_i commute. In the latter case, the *upper* bound of (1.2) is typically of the correct order. Indeed, by simultaneously diagonalizing A_i, we may assume X is a diagonal matrix. Then $\sigma(X)^2 = \|\mathbf{E}X^2\| = \max_i \mathrm{Var}(X_{ii})$, while

$$\mathbf{E}\|X\| = \mathbf{E}\max_i |X_{ii}| \asymp \sigma(X)\sqrt{\log d}$$

under mild assumptions (as the maximum of d Gaussian variables is typically of order $\sqrt{\log d}$, see, e.g., [24, §3.3]). On the other hand, when the coefficients A_i do

[1] We write $x \lesssim y$ if $x \leq Cy$ for a universal constant C, and $x \asymp y$ if $x \lesssim y$ and $y \lesssim x$.

not commute, it is observed in many examples that it is the *lower* bound of (1.2) that is of the correct order. This is already the case for the most basic model of random matrix theory: when X has i.i.d. standard Gaussian entries X_{ij} for $i \geq j$, it is classical that $\mathbf{E}\|X\| \asymp \sqrt{d} = \sigma(X)$ [33, §2.3].

Such examples raise the tantalizing question whether there exists a refinement of (1.2) that can capture the correct behavior of nonhomogeneous random matrices beyond the commutative case. To date, a satisfactory answer to this question has been obtained only in the special case that X has *independent* entries X_{ij} for $i \geq j$ with arbitrary variances $\mathrm{Var}(X_{ij}) = b_{ij}^2$. In this case, [5] showed that

$$\mathbf{E}\|X\| \lesssim \sigma(X) + \max_{ij} |b_{ij}| \sqrt{\log d}, \qquad \sigma(X)^2 = \max_i \sum_j b_{ij}^2, \tag{1.4}$$

which can be reversed under mild assumptions. The key feature of (1.4) is that the dimensional factor enters here through a smaller parameter $\sigma_*(X) = \max_{ij} |b_{ij}|$ that controls which extreme case of (1.2) dominates: diagonal matrices satisfy $\sigma_*(X) = \sigma(X)$, in which case we recover the upper bound of (1.2); but as soon as $\sigma_*(X) \lesssim (\log d)^{-\frac{1}{2}}\sigma(X)$, the lower bound of (1.2) is of the correct order.

The existence of the bound (1.4) hints at the possibility that an analogous refinement of (1.2) might hold even in the setting of general Gaussian random matrices (1.1). In particular, one may conjecture the existence of a general bound

$$\mathbf{E}\|X\| \overset{?}{\lesssim} \sigma(X) + \sigma_{**}(X)(\log d)^\beta \tag{1.5}$$

for some $\beta > 0$, where $\sigma(X)$ is as in (1.3) and $\sigma_{**}(X)$ is a parameter that is small when the coefficients A_i are far from being commutative. This question was first considered by Tropp [37], who introduced a number of important ideas that form the basis for the present paper. Using these ideas, Tropp was able to prove a bound of the form (1.5) for a special class of models that satisfy strong symmetry assumptions (and for general models with a dimensional factor $(\log d)^{\frac{1}{4}}$ in the leading term). To date, however, a general bound of the form (1.5) has remained elusive.

1.2. **Free probability.** The challenge in proving an inequality of the form (1.5) is to capture the intrinsic noncommutativity of the matrices A_i. There is however an entirely different way to introduce noncommutativity into (1.1) that arises from Voiculescu's theory of free probability [40, 27]: one may modify the model by replacing the scalar Gaussian coefficients g_i by noncommuting random matrices or operators. When noncommutativity is externally introduced into (1.1) in this manner, the dimensional factor in (1.2) is unnecessary regardless of the properties of the matrices A_i (see (1.10) below). However, on its face, this appears to shed little light on the behavior of the original model (1.1).

Remarkably, this intuition proves to be incorrect. The central theme that will be developed in this paper is described informally by the following principle:

> *When the coefficient matrices A_i are sufficiently noncommutative, the spectral statistics of the random matrix model $X = \sum_i g_i A_i$ are already accurately captured by free probability.*

This "intrinsic freeness" phenomenon will prove to have far-reaching implications: it will enable us to prove nonasymptotic bounds of the form (1.5) in complete generality (both for Gaussian random matrices and for general sums of independent random matrices), and to develop new asymptotic results in free probability in far more general situations than are accessible by previous methods.

Before we can formulate precise results along these lines, we must briefly recall some relevant notions of free probability. We will use the following terminology.

Definition 1.1. A *standard Wigner matrix* of dimension N is an $N\times N$ self-adjoint random matrix G^N whose entries on and above the diagonal are independent real Gaussian variables with mean zero and variance $\frac{1}{N}$.

Free probability provides an asymptotic description of the behavior of Wigner matrices as $N\to\infty$. Let $G_1^N,\ldots,G_n^N$ be independent standard Wigner matrices; the associated limiting objects are certain infinite-dimensional self-adjoint operators $s_1,\ldots,s_n$ that form a *free semicircular family*, together with a trace τ acting on the algebra generated by these operators. We postpone the precise definitions of these objects to Section 4.1; for our purposes, they may be viewed as an algebraic tool that allows us to compute spectral properties of large random matrices. In particular, a celebrated result of Voiculescu [40] states that

$$\lim_{N\to\infty}\mathbf{E}[\operatorname{tr} p(G_1^N,\ldots,G_n^N)] = \tau(p(s_1,\ldots,s_n)) \tag{1.6}$$

for any noncommutative polynomial p, where $\operatorname{tr}(M) := \frac{1}{N}\operatorname{Tr}(M)$ denotes the normalized trace of a matrix $M\in\mathrm{M}_N(\mathbb{C})$. In an important paper, Haagerup and Thorbjørnsen [19] showed that the *weak asymptotic freeness* property (1.6) may be considerably strengthened to obtain convergence in norm

$$\lim_{N\to\infty}\mathbf{E}[\|p(G_1^N,\ldots,G_n^N)\|] = \|p(s_1,\ldots,s_n)\| \tag{1.7}$$

for any noncommutative polynomial p. This *strong asymptotic freeness* property has important applications in the theory of operator algebras [19, 17, 18].

A noncommutative analogue of the random matrix model (1.1) is obtained by replacing the scalar Gaussian coefficients g_i by standard Wigner matrices:

$$X^N = \sum_{i=1}^n A_i\otimes G_i^N. \tag{1.8}$$

When $N=1$, this model coincides with (1.1); however, as N increases, the matrices G_i^N become increasingly noncommutative. The weak and strong asymptotic freeness properties (1.6) and (1.7) imply that the behavior of the spectrum of X^N as $N\to\infty$ is captured by the infinite-dimensional operator

$$X_{\mathrm{free}} = \sum_{i=1}^n A_i\otimes s_i \tag{1.9}$$

in that $\lim_{N\to\infty}\mathbf{E}\operatorname{tr}[(X^N)^p] = (\operatorname{tr}\otimes\tau)(X_{\mathrm{free}}^p)$ and $\lim_{N\to\infty}\mathbf{E}\|X^N\| = \|X_{\mathrm{free}}\|$. The study of such models plays a fundamental role in [19].

While X_{free} may be viewed abstractly as the limiting object associated to X^N, its considerable utility (from the perspective of this paper) is that it enables explicit computation of many spectral statistics of the random matrices X^N. For example, as we will recall in Section 2.1, the norm $\|X_{\mathrm{free}}\|$ admits an explicit formula in terms of the matrices A_i [25] and admits the simple estimates [28, p. 208]

$$\sigma(X)\le\|X_{\mathrm{free}}\|\le 2\sigma(X). \tag{1.10}$$

Similarly, the limiting spectral distribution of X^N may be computed by means of a (matrix-valued) Dyson equation as in classical random matrix theory [19, 1].

1.3. **Main results.** We now give an brief overview of the main results of this paper. A detailed presentation of our main results will be given in Section 2, while various examples that illustrate our results will be discussed in Section 3.

1.3.1. *Gaussian random matrices.* To illustrate the general principle described in Section 1.2, let us begin by stating a special case of one of our main results. For any centered $d \times d$ random matrix X as in (1.1), we denote by $\mathrm{Cov}(X) \in M_{d^2}(\mathbb{C})_{\mathrm{sa}}$ the covariance matrix of its d^2 scalar entries, that is,

$$\mathrm{Cov}(X)_{ij,kl} = \mathbf{E}[X_{ij}\overline{X_{kl}}] = \sum_{s=1}^{n} (A_s)_{ij}\overline{(A_s)_{kl}}$$

which we view as a $d^2 \times d^2$ positive semidefinite matrix. We now define

$$v(X)^2 = \|\mathrm{Cov}(X)\| = \sup_{\mathrm{Tr}|M|^2 \le 1} \sum_{s=1}^{n} |\mathrm{Tr}[A_s M]|^2.$$

It should be far from apparent at this point that the parameter $v(X)$ captures noncommutativity of the matrices A_i; this will be explained in Section 1.4. Note, for example, that $v(X) \asymp \max_{ij} |b_{ij}|$ in the setting of (1.4) (cf. section 3.1).

Theorem 1.2. *For the model* (1.1) *we have*

$$\mathbf{E}\|X\| \le \|X_{\mathrm{free}}\| + C\, v(X)^{\frac{1}{2}} \sigma(X)^{\frac{1}{2}} (\log d)^{\frac{3}{4}},$$

where X_{free} *is defined in* (1.9) *and* C *is a universal constant.*

Using (1.10) and Young's inequality, Theorem 1.2 immediately implies a completely general bound of the form (1.5):

$$\mathbf{E}\|X\| \lesssim \sigma(X) + v(X)(\log d)^{\frac{3}{2}}. \tag{1.11}$$

However, Theorem 1.2 is much sharper in that its leading term captures the exact quantity predicted by free probability. In many cases, our results will make it possible to prove that $\mathbf{E}\|X\| = (1+o(1))\|X_{\mathrm{free}}\|$ as soon as $v(X)/\sigma(X) = o((\log d)^{-\frac{3}{2}})$, providing essentially optimal results in this setting.

Our main results for Gaussian random matrices (see Sections 2.1 and 2.2) are considerably more general than Theorem 1.2. In particular:

- Our main results are formulated for arbitrary Gaussian random matrices, which may have nonzero mean and may be non-self-adjoint.
- We bound the support of the full spectrum $\mathrm{sp}(X) \subseteq \mathrm{sp}(X_{\mathrm{free}}) + [-\varepsilon, \varepsilon]$ with high probability, where $\varepsilon \asymp v(X)^{\frac{1}{2}} \sigma(X)^{\frac{1}{2}} (\log d)^{\frac{3}{4}}$.
- We obtain nonasymptotic upper and lower bounds on the moments, resolvent, and other spectral statistics of X in terms of X_{free}.

The "intrinsic freeness" phenomenon that is captured by these results has strong implications both for matrix concentation inequalities and in free probability.

1.3.2. *Asymptotic freeness.* While our main results are nonasymptotic in nature, they give rise to remarkable new asymptotic results in free probability: when combined with the linearization trick of [19], our results establish strong asymptotic freeness (1.7) for a very large class of random matrix models. For example, we will prove the following result, as well as an analogous strong law (which yields a.s. convergence) that will be formulated in Section 2.3.

Theorem 1.3. *Let $s_1, \ldots, s_m$ be a free semicircular family. For each $N \geq 1$, let $H_1^N, \ldots, H_m^N$ be independent self-adjoint random matrices of dimension $d = d(N) \geq N$ such that each H_k^N has jointly Gaussian entries, $\mathbf{E}[H_k^N] = 0$, and $\mathbf{E}[(H_k^N)^2] = \mathbf{1}$.*

a. If $v(H_k^N) = o(1)$ as $N \to \infty$ for all k, then for any polynomial p

$$\lim_{N\to\infty} \mathbf{E}[\operatorname{tr} p(H_1^N, \ldots, H_m^N)] = \tau(p(s_1, \ldots, s_m)).$$

b. If $v(H_k^N) = o((\log d)^{-\frac{3}{2}})$ as $N \to \infty$ for all k, then for any polynomial p

$$\lim_{N\to\infty} \mathbf{E}[\|p(H_1^N, \ldots, H_m^N)\|] = \|p(s_1, \ldots, s_m)\|.$$

A striking consequence of Theorem 1.3 is the unexpected ubiquity of the strong asymptotic freeness property. To date, strong asymptotic freeness has been proved only for Wigner matrices and for certain highly symmetric ensembles; for a detailed overview of prior results, see [10, 6] and the references cited therein. In contrast, neither symmetry nor independent entries plays any role in Theorem 1.3, which enables us to establish strong asymptotic freeness in models that appear to lie far outside the reach of previous methods (for example, for sparse Wigner matrices of dimension d with only $O(d \log^4 d)$ nonzero entries, see Example 3.5). For many such models, even weak asymptotic freeness (1.6) was not previously known.

1.3.3. *Sums of independent random matrices.* When viewed as matrix concentration inequalities, bounds such as (1.11) are easily applicable in concrete situations and yield results of optimal order in many examples where classical matrix concentration inequalities are suboptimal, as will be illustrated in Section 3.

While our sharpest results are obtained in the Gaussian setting, the Gaussian assumption may be restrictive in applications (particularly in applied mathematics). However, using routine arguments, we may extend our bounds to a much more general setting at the expense of losing a universal constant. For example, we will derive the following result for arbitrary sums of independent random matrices.

Theorem 1.4. *Let $Z_1, \ldots, Z_n$ be arbitrary independent $d \times d$ self-adjoint centered random matrices, and let $X = \sum_{i=1}^n Z_i$. Then*

$$\mathbf{E}\|X\| \lesssim \|\mathbf{E}X^2\|^{\frac{1}{2}} + \|\mathrm{Cov}(X)\|^{\frac{1}{2}} (\log d)^{\frac{3}{2}} + (\mathbf{E}[\max_i \|Z_i\|_{\mathrm{HS}}^2])^{\frac{1}{2}} (\log d)^2,$$

where $\|Z\|_{\mathrm{HS}}^2 := \mathrm{Tr}[|Z|^2]$ denotes the Hilbert-Schmidt norm.

Theorem 1.4 yields considerably stronger results than the widely used matrix Bernstein inequality [35] in various situations. An extension of Theorem 1.4 to non-symmetric matrices and a corresponding tail bound is given in Section 2.4.

1.4. **Overview of the proofs.**

1.4.1. *Crossings.* Before we describe the main technique used in our proofs, let us briefly outline the origin of the key parameter $v(X)$ that quantifies noncommutativity in our results, and its relation to free probability.

The simplest way to understand the difference between the random matrix X and its free counterpart X_{free} is in terms of their moments. Let us recall that these moments may be expressed combinatorially as [27, pp. 128–129]

$$\mathbf{E}[\operatorname{tr} X^{2p}] = \sum_{\pi \in \mathrm{P}_2([2p])} \sum_{(i_1, \ldots, i_{2p}) \sim \pi} \operatorname{tr}[A_{i_1} \ldots A_{i_{2p}}]$$

and

$$(\mathrm{tr}\otimes\tau)(X_{\mathrm{free}}^{2p}) = \sum_{\pi\in\mathrm{NC}_2([2p])}\ \sum_{(i_1,\ldots,i_{2p})\sim\pi} \mathrm{tr}[A_{i_1}\ldots A_{i_{2p}}],$$

where $\mathrm{P}_2([2p])$ and $\mathrm{NC}_2([2p])$ denote the families of all pair partitions and noncrossing pair partitions of $[2p]$, respectively, and $(i_1,\ldots,i_{2p}) \sim \pi$ signifies that $i_k = i_l$ whenever $\{k,l\}\in\pi$. In other words, what distinguishes free probability from classical probability is the absence of crossings, that is, of terms of the form $\sum_{ij}\cdots A_i\cdots A_j\cdots A_i\cdots A_j\cdots$ in the moment formulae.

In free probability, the vanishing of crossings arises from the noncommutativity of the semicircular family s_i. Even in (1.1), however, crossings may be intrinsically suppressed due to the noncommutativity of the coefficients A_i. It is a beautiful idea of Tropp [37] to quantify the latter effect by the parameter

$$w(X) = \sup_{U,V,W} \|\mathbf{E}[X_1UX_2VX_1WX_2]\|^{\frac{1}{4}} = \sup_{U,V,W} \left\| \sum_{i,j=1}^n A_iUA_jVA_iWA_j \right\|^{\frac{1}{4}},$$

where X_1,X_2 are i.i.d. copies of X and the supremum is taken over all (nonrandom) unitary matrices U,V,W of the same dimension as X. Note that when all A_i commute, $w(X) \geq \|\sum_{ij}A_iA_jA_iA_j\|^{\frac{1}{4}} = \|(\sum_i A_i^2)^2\|^{\frac{1}{4}} = \sigma(X)$; but if $w(X) \ll \sigma(X)$, the contribution of crossings will be suppressed.

Unfortunately, the quantity $w(X)$ is very unwieldy and is difficult to compute in practice. Moreover, as will be explained below, the quantity that will arise in our proofs is not $w(X)$, but rather $w(\tilde{X})$ for an auxiliary matrix $\tilde{X}$ of much higher dimension. To control this parameter, we will show in Section 4.2 that

$$w(X) \leq v(X)^{\frac{1}{2}}\sigma(X)^{\frac{1}{2}}, \tag{1.12}$$

which enables us to formulate our results in terms of the much simpler quantity $v(X)$ that is readily computable in concrete situations. In particular, it follows that $v(X)$ does indeed capture noncommutativity, as it controls $w(X)$.

The notion that smallness of $w(X)$ should lead to free behavior is implicit in the work of Tropp [37]. However, the attempt in [37] to exploit this idea by means of moment recursions appears to be insufficiently powerful to capture this phenomenon without imposing strong symmetry assumptions on the coefficients A_i. A key new idea of this paper enables us to capture this phenomenon in its full strength.

1.4.2. *Interpolation.* The central idea behind our proofs is the following construction. Let $G_1^N,\ldots,G_n^N$ be independent standard Wigner matrices as in Section 1.2, and let $D_1^N,\ldots,D_n^N$ be independent $N\times N$ diagonal matrices with i.i.d. standard Gaussians on the diagonal. Define for $q\in[0,1]$ the random matrix

$$X_q^N = \sum_{i=1}^n A_i\otimes(\sqrt{q}\,D_i^N + \sqrt{1-q}\,G_i^N).$$

The point of this construction is that the family $(X_q^N)_{q\in[0,1],N\in\mathbb{N}}$ enables us to interpolate between X and X_{free}. Indeed, $X_0^N = X^N$ is the model (1.8), whose moments converge as $N\to\infty$ to those of X_{free} by (1.6) (this is the only property that will be used in our proofs; strong asymptotic freeness will not be assumed). On the other hand, it is readily verified that X_1^N has the same moments as X in the sense $\mathbf{E}[\mathrm{tr}\,X^p] = \mathbf{E}\,\mathrm{tr}[(X_1^N)^p]$ for every $p,N\in\mathbb{N}$.

In order to bound the moments of X by those of X_{free}, it suffices to bound the rate at which the moments change along the above interpolation. Given that the moments of X and X_{free} differ only by terms involving crossings, it is natural to expect that the rate of change along the interpolation will be controlled by the contributions of the crossings. It will turn out that the construction of the matrices X_q^N has precisely the right form in order to capture this phenomenon in terms of the parameters described in the previous section. More precisely, the explicit expression for the derivative $\frac{d}{dq}\mathbf{E}\,\mathrm{tr}[(X_q^N)^p]$, which can be computed using a standard Gaussian interpolation lemma [32, §1.3], can be controlled in terms of the quantity

$$\tilde{w}(X) = \sup_N w(X_1^N).$$

The resulting differential inequality may be integrated to bound the moments of X in terms of the moments of X_{free} and the parameter $\tilde{w}(X)$. As the latter is nearly impossible to compute, we finally obtain a practical bound $\tilde{w}(X) \le v(X)^{\frac{1}{2}}\sigma(X)^{\frac{1}{2}}$ using (1.12) and $v(X_1^N) = v(X)$, $\sigma(X_1^N) = \sigma(X)$.

The above interpolation method proves to be a powerful tool for capturing "intrinsic freeness". The same method can be used to control not just the moments, but also various other spectral statistics. In particular, we will control the full spectrum of X by that of X_{free} by applying the interpolation method to large moments of the resolvent $\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X|^{-2p}]$. Such control of the full spectrum is crucial for the applications of our results to free probability described in Section 1.3.2.

Remark 1.5. After the results of this paper were completed, we learned that a different interpolation method was recently used by Collins, Guionnet, and Parraud [10] to obtain a quantitative form of the strong asymptotic freeness of Wigner matrices due to Haagerup-Thorbjørnsen. Rather than interpolating between scalar Gaussians and Wigner matrices, [10] interpolate in the opposite direction, between Wigner matrices and a semicircular family, using the free Ornstein-Uhlenbeck semigroup. The latter does not capture any notion of "intrinsic freeness". Nonetheless, the results of [10] and of the present paper suggest that interpolation provides a versatile method to obtain quantitative results in free probability.

1.5. **Organization of this paper.** The rest of this paper is organized as follows. In Section 2, the main results of this paper will be presented in full detail. The utility of our main results will then be illustrated in a number of concrete examples in Section 3. Section 4 briefly reviews some basic notions of free probability, and introduces various tools that are used throughout the rest of the paper. The proofs of our main results are given in Sections 5–8.

The final Section 9 is devoted to a discussion of various broader questions arising from our main results. In particular, we will show that there cannot exist a canonical choice of the parameter $\sigma_{**}(X)$ in the inequality (1.5), as any such parameter must violate some natural property of the spectral norm. This disproves a conjecture, formulated in [37, 39, 4], which suggests that the parameter $v(X)$ in our main results can be replaced by a certain smaller parameter $\sigma_*(X)$ that will be defined below. We conclude by discussing a number of open questions.

1.6. **Notation.** The following notations will be frequently used throughout this paper. We write $[n] := \{1, \ldots, n\}$ for $n \in \mathbb{N}$. For a bounded operator X on a Hilbert space, we denote by $\|X\|$ its operator (i.e., spectral) norm and by $|X| := (X^*X)^{\frac{1}{2}}$. The spectrum of X is denoted as $\mathrm{sp}(X)$. If X is self-adjoint and $h : \mathbb{R} \to \mathbb{C}$ is measurable, then the operator $h(X)$ is defined by the usual functional calculus (in particular, if X is a self-adjoint matrix, h is applied to the eigenvalues while keeping the eigenvectors fixed). The algebra of $d\times d$ matrices with values in a *-algebra $\mathcal{A}$ is denoted as $\mathrm{M}_d(\mathcal{A})$, and its subspace of self-adjoint matrices is denoted as $\mathrm{M}_d(\mathcal{A})_{\mathrm{sa}}$. For complex matrices $M \in \mathrm{M}_d(\mathbb{C})$, we always denote by $\mathrm{Tr}\, M := \sum_{i=1}^d M_{ii}$ the unnormalized trace and by $\mathrm{tr}\, M := \frac{1}{d}\,\mathrm{Tr}\, M$ the normalized trace.

2. Main results

2.1. Concentration of the spectrum. The strongest results of this paper apply to arbitrary random matrices with jointly Gaussian entries (this model is more general than the one that was assumed for sake of illustration in the introduction). To define this model, fix $d \ge 2$ and $n \in \mathbb{N}$, let $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})$, let $g_1,\ldots,g_n$ be i.i.d. real Gaussian variables with zero mean and unit variance, and let $s_1,\ldots,s_n$ be a free semicircular family (cf. Section 4.1). We now define

$$X := A_0 + \sum_{i=1}^n g_i A_i, \qquad X_{\mathrm{free}} := A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes s_i. \tag{2.1}$$

In formulating our results, it will sometimes be convenient to assume in addition that the model is self-adjoint, that is, that $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$. In such cases this assumption will be made merely for notational convenience and is not a restriction, as will be explained in Remark 2.6 below.

The following parameters will play a fundamental role in the sequel:

$$\begin{aligned}
\sigma(X)^2 &:= \left\| \sum_{i=1}^n A_i^* A_i \right\| \vee \left\| \sum_{i=1}^n A_i A_i^* \right\| = \|\mathbf{E}\hat{X}^*\hat{X}\| \vee \|\mathbf{E}\hat{X}\hat{X}^*\|,\\
\sigma_*(X)^2 &:= \sup_{\|v\|=\|w\|=1} \sum_{i=1}^n |\langle v, A_i w\rangle|^2 = \sup_{\|v\|=\|w\|=1} \mathbf{E}[|\langle v, \hat{X} w\rangle|^2],\\
v(X)^2 &:= \sup_{\mathrm{Tr}|M|^2\le 1} \sum_{i=1}^n |\mathrm{Tr}[A_i M]|^2 = \|\mathrm{Cov}(X)\|,
\end{aligned}$$

where $\hat{X} := X - \mathbf{E}X$. It follows readily from the definitions that $\sigma_*(X) \le v(X)$ and $\sigma_*(X) \le \sigma(X)$. As the following combination will appear frequently, we let

$$\tilde{v}(X)^2 := v(X)\sigma(X).$$

Note that the definitions of these parameters do not involve A_0.

We can now formulate our main result on concentration of the spectrum of X. Here $\mathrm{sp}(M)$ denotes the spectrum of a self-adjoint operator M.

Theorem 2.1. *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$, *we have*

$$\mathbf{P}\big[\mathrm{sp}(X) \subseteq \mathrm{sp}(X_{\mathrm{free}}) + C\{\tilde{v}(X)(\log d)^{\frac{3}{4}} + \sigma_*(X)t\}[-1,1]\big] \ge 1 - e^{-t^2}$$

for all $t \ge 0$, *where* C *is a universal constant.*

The spectrum of X_{free} always consists of a finite union of bounded intervals [1]. Theorem 2.1 implies that when $v(X) \ll (\log d)^{-\frac{3}{2}}\sigma(X)$, *all* eigenvalues of X are close to the spectrum of X_{free}. In particular, not only must the extreme eigenvalues of X lie close to the edge of the spectrum of X_{free}, but also the interior eigenvalues cannot lie far inside the gaps in the spectrum of X_{free}.

When specialized to the extreme eigenvalues, Theorem 2.1 yields a bound on the spectral norm of X. We formulate it here directly for non-self-adjoint matrices.

Corollary 2.2. *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})$, *we have*

$$\mathbf{P}\big[\|X\| > \|X_{\mathrm{free}}\| + C\tilde{v}(X)(\log d)^{\frac{3}{4}} + C\sigma_*(X)t\big] \le e^{-t^2}$$

for all $t \ge 0$, *where* C *is a universal constant. Moreover,*

$$\mathbf{E}\|X\| \le \|X_{\mathrm{free}}\| + C\tilde{v}(X)(\log d)^{\frac{3}{4}}.$$

Theorem 2.1 and Corollary 2.2 will be proved in Section 6.

Remark 2.3. In order to apply Corollary 2.2 in concrete situations, we must be able to compute or estimate $\|X_{\mathrm{free}}\|$. For ease of reference, we presently recall two useful facts; further discussion and references may be found in Section 4.1. In the following, $\lambda_{\max}(M)$ denotes the maximal eigenvalue of a self-adjoint matrix M.

Lemma 2.4 (Lehner)**.** *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$, *we have*

$$\|X_{\mathrm{free}}\| = \max_{\varepsilon=\pm 1} \inf_{Z>0} \lambda_{\max}\left(Z^{-1} + \varepsilon A_0 + \sum_{i=1}^n A_i Z A_i\right),$$

where the infimum is over positive definite $Z \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$. *The infimum may be further restricted to* Z *for which the matrix in* $\lambda_{\max}(\cdots)$ *is a multiple of the identity.*

Lemma 2.5 (Pisier)**.** *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})$, *we have*

$$\|A_0\| \vee \sigma(X) \le \|X_{\mathrm{free}}\| \le \|A_0\| + \left\|\sum_{i=1}^n A_i^* A_i\right\|^{\frac{1}{2}} + \left\|\sum_{i=1}^n A_i A_i^*\right\|^{\frac{1}{2}}.$$

Note that the combination of Corollary 2.2 and Lemma 2.5 immediately yields a Gaussian matrix concentration inequality of the form (1.5).

Remark 2.6. For simplicity, we formulated results such as Theorem 2.1 and Lemma 2.4 for self-adjoint matrices. The following standard device makes it possible to reduce the general case to the self-adjoint case. Given $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})$, define the matrices $\breve{A}_0,\ldots,\breve{A}_n \in M_{2d}(\mathbb{C})_{\mathrm{sa}}$, $\breve{X}$, and $\breve{X}_{\mathrm{free}}$ as

$$\breve{A}_i = \begin{bmatrix} 0 & A_i \\ A_i^* & 0 \end{bmatrix}, \qquad \breve{X} = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}, \qquad \breve{X}_{\mathrm{free}} = \begin{bmatrix} 0 & X_{\mathrm{free}} \\ X_{\mathrm{free}}^* & 0 \end{bmatrix}.$$

Then it is not difficult to show (see Section 4.2.3) that

$$\mathrm{sp}(\breve{X}) \cup \{0\} = \mathrm{sp}(|X|) \cup -\mathrm{sp}(|X|) \cup \{0\},$$

and analogously for X_{free}; moreover, we have

$$\sigma(\breve{X}) = \sigma(X), \qquad \sigma_*(\breve{X}) = \sigma_*(X), \qquad v(\breve{X}) \le \sqrt{2}\,v(X).$$

Applying Theorem 2.1 to $\breve{X}$ therefore shows that in the non-self-adjoint case, the *singular values* of X concentrate around those of X_{free}. Similarly, we can apply Lemma 2.4 to $\breve{X}_{\mathrm{free}}$ to obtain an explicit formula for $\|X_{\mathrm{free}}\|$.

The above construction does not require the matrices A_i to be square. However, if A_i are $d_1 \times d_2$ matrices with $d_1 < d_2$, the singular values of X are unchanged if we add $d_2 - d_1$ zero rows to the matrix. Thus there is no loss of generality in restricting attention to square matrices, as we do for simplicity throughout this paper.

2.2. **Spectral statistics.** The results of the previous section quantify concentration of the eigenvalues of X near the spectrum of $X_{\rm free}$. We now formulate several complementary results that quantify the closeness of the spectral distributions of X and $X_{\rm free}$. We begin by stating a bound on the moments.

Theorem 2.7. *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})$*, we have*

$$|\mathbf{E}[\mathrm{tr}\,|X|^{2p}]^{\frac{1}{2p}} - (\mathrm{tr}\otimes\tau)(|X_{\rm free}|^{2p})^{\frac{1}{2p}}| \le 2p^{\frac{3}{4}}\tilde{v}(X)$$

for all $p\in\mathbb{N}$.

Let us emphasize that unlike the results of Section 2.1, Theorem 2.7 gives a two-sided bound on X in terms of $X_{\rm free}$. This opens the door to obtaining sharp asymptotics from our nonasymptotic bounds.

The same method of proof is readily applied to other spectral statistics. To illustrate this, we will bound the matrix-valued Stieltjes transform, which plays an important role in operator-valued free probability [26, Chapters 9–10]. A bound of this kind is most naturally formulated for self-adjoint matrices.

Theorem 2.8. *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})_{\rm sa}$*, define the matrix-valued Stieltjes transforms* $G(Z), G_{\rm free}(Z)\in\mathrm{M}_d(\mathbb{C})$ *as*

$$G(Z) := \mathbf{E}[(Z-X)^{-1}], \qquad G_{\rm free}(Z) := (\mathrm{id}\otimes\tau)[(Z\otimes\mathbf{1}-X_{\rm free})^{-1}].$$

Then we have

$$\|G(Z)-G_{\rm free}(Z)\| \le \tilde{v}(X)^4\|(\mathrm{Im}\,Z)^{-5}\|$$

for all $Z\in\mathrm{M}_d(\mathbb{C})$ *with* $\mathrm{Im}\,Z := \frac{1}{2i}(Z-Z^*)>0$.

Following [19, §6], Theorem 2.8 implies a bound on smooth spectral statistics.

Corollary 2.9. *For the model* (2.1) *with* $A_0,\ldots,A_n \in \mathrm{M}_d(\mathbb{C})_{\rm sa}$*, we have*

$$|\mathbf{E}[\mathrm{tr}\,f(X)] - (\mathrm{tr}\otimes\tau)[f(X_{\rm free})]| \lesssim \tilde{v}(X)^4\|f\|_{W^{6,1}(\mathbb{R})}$$

for every $f\in W^{6,1}(\mathbb{R})$.

Theorems 2.7–2.8 and Corollary 2.9 will be proved in Section 5.

2.3. **Strong asymptotic freeness.** By combining the bounds of Sections 2.1–2.2 with the linearization trick of [19], we will be able to establish strong asymptotic freeness for a remarkably general class of random matrices. We presently give a complete formulation of our main result in this direction.

Theorem 2.10. *Let* $s_1,\ldots,s_m$ *be a free semicircular family. For each* $N\ge1$*, let* $H_1^N,\ldots,H_m^N$ *be independent self-adjoint random matrices of dimension* $d=d(N)\ge N$ *such that each* H_k^N *has jointly Gaussian entries,*

$$\lim_{N\to\infty}\|\mathbf{E}[H_k^N]\|=0, \qquad \lim_{N\to\infty}\|\mathbf{E}[(H_k^N)^2]-\mathbf{1}\|=0$$

for all k*. Then the following hold.*

a. If $v(H_k^N)=o(1)$ *as* $N\to\infty$ *for all* k*, then*

$$\lim_{N\to\infty}\mathbf{E}[\mathrm{tr}\,p(H_1^N,\ldots,H_m^N)] = \tau(p(s_1,\ldots,s_m))$$

for every noncommutative polynomial p.

b. If $v(H_k^N) = o((\log d)^{-\frac{3}{2}})$ as $N \to \infty$ for all k, then

$$\lim_{N\to\infty} \mathbf{E}[\|p(H_1^N,\ldots,H_m^N)\|] = \|p(s_1,\ldots,s_m)\|,$$
$$\lim_{N\to\infty} \|p(H_1^N,\ldots,H_m^N)\| = \|p(s_1,\ldots,s_m)\| \quad a.s.,$$
$$\lim_{N\to\infty} \operatorname{tr} p(H_1^N,\ldots,H_m^N) = \tau(p(s_1,\ldots,s_m)) \quad a.s.$$

for every noncommutative polynomial p.

Let us recall that the type of convergence in part *b* of Theorem 2.10, called *strong convergence in distribution*, has even stronger implications: it implies that both the spectral distribution and support of the spectrum of any polynomial $p(H_1^N,\ldots,H_m^N)$ converges to that of $p(s_1,\ldots,s_m)$ as $N\to\infty$ in the sense of weak convergence and Hausdorff convergence, respectively; see [11, Proposition 2.1].

Surprisingly, the conclusion of Theorem 2.10 appears to be new at this level of generality already for a single random matrix $m = 1$. In this case, we obtain the following result in the spirit of classical random matrix theory.

Corollary 2.11. *Let H^N be a self-adjoint random matrix of dimension $d = d(N)$ with jointly Gaussian entries, and assume that*

$$\|\mathbf{E}[H^N]\| = o(1), \qquad \|\mathbf{E}[(H^N)^2] - \mathbf{1}\| = o(1), \qquad v(H^N) = o((\log d)^{-\frac{3}{2}})$$

as $N \to \infty$. Then the empirical distribution

$$\mu_{H^N} := \frac{1}{d}\sum_{i=1}^{d} \delta_{\lambda_i(H^N)}$$

of the eigenvalues $\lambda_i(H^N)$ of H^N converges weakly a.s. to the semicircle law

$$\mu_{H^N} \xrightarrow{\text{w}} \mu_{\mathrm{sc}} \quad a.s., \qquad \mu_{\mathrm{sc}}(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\, 1_{|x|\le 2}\, dx,$$

and we have convergence of the norm $\|H^N\| \to 2$ a.s. as $N \to \infty$.

Let us emphasize that Corollary 2.11 (and Theorem 2.10) makes no structural assumptions on the variance or dependence pattern of H^N beyond the minimal isotropy conditions $\mathbf{E}[H^N] \approx 0$ and $\mathbf{E}[(H^N)^2] \approx \mathbf{1}$. Previous results on Gaussian random matrices with dependent entries require restrictive structural assumptions to obtain even the semicircle law, cf. [14] and the references therein.

Theorem 2.10 and Corollary 2.11 will be proved in Section 7.

2.4. **Matrix concentration inequalities.** All the results presented above apply to Gaussian random matrices. However, using routine symmetrization arguments, we can extend our bounds to a much more general class of random matrices at the expense of the loss of a universal constant. Because of the latter, the resulting bounds can no longer capture the exact free probability behavior that is a central feature of our Gaussian results. Nonetheless, such bounds can be very useful in many applications due to their generality and simplicity. We presently formulate a concrete result along these lines, which will be proved in Section 8.

Theorem 2.12. *Let $Z_1,\ldots,Z_n$ be arbitrary independent $d\times d$ random matrices with $\mathbf{E}[Z_i] = 0$, and let $X = \sum_{i=1}^n Z_i$. Define the matrix parameters*

$$v^2 = \|\mathrm{Cov}(X)\|, \qquad L^2 = \mathbf{E}[\max_i \|Z_i\|_{\mathrm{HS}}^2], \qquad \sigma_*^2 = \sup_{\|v\|=\|w\|=1} \mathbf{E}[|\langle v, Xw\rangle|^2],$$

and

$$\sigma_1^2 = \|\mathbf{E}X^*X\|, \qquad \sigma_2^2 = \|\mathbf{E}XX^*\|.$$

Then

$$\mathbf{E}\|X\| \le 2\{\sigma_1+\sigma_2\} + C(\log d)^{\frac{3}{2}}v + C(\log d)^2 L$$

for a universal constant C. Moreover, if $\|Z_i\| \le R$ a.s. for all i, then

$$\mathbf{P}[\|X\| \ge 2\{\sigma_1+\sigma_2\} + C\{(\log d)^{\frac{3}{2}}v + (\log d)^2 L + \sigma_*\sqrt{t} + Rt\}] \le e^{-t}$$

for all $t \ge 0$.

Note that we have assumed in Theorem 2.12 that $\mathbf{E}[X] = 0$. For sums of independent random matrices with nonzero mean, Theorem 2.12 may be applied to the random matrix $X - \mathbf{E}[X]$; thus Theorem 2.12 may indeed be interpreted as a "matrix concentration inequality," in the sense that it bounds the deviation of a sum of independent random matrices from its mean (in the spectral norm). The present bound should be compared with the widely used matrix Bernstein inequality [35, Theorem 6.1.1], which states that

$$\mathbf{E}\|X\| \lesssim (\sigma_1+\sigma_2)\sqrt{\log d} + R\log d. \tag{2.2}$$

Theorem 2.12 replaces the dimensional factor in the leading term of the matrix Bernstein inequality by a constant factor 2 when v, L, R are sufficiently small, which is the case in many examples (e.g., for random matrices with bounded i.i.d. entries). Let us emphasize that while the leading term of Theorem 2.12 is dimension-free, the constant 2 is not optimal; see section 9.2.2 for further discussion.

Remark 2.13. As $\|\mathrm{Cov}(X)\| = \sup_{\|M\|_{\mathrm{HS}}\le 1} \mathbf{E}[|\mathrm{Tr}[MX]|^2]$, we have $\sigma_* \le v$ in Theorem 2.12. The parameter v is often easier to compute in practice.

Remark 2.14. Note that the parameters $\sigma_1, \sigma_2, \sigma_*, v$ in Theorem 2.12 depend only on the covariance of the entries of X, while the parameters L, R do not. One may think of L, R as parameters that bound the deviation from Gaussianity. For example, suppose $Y_1, \ldots, Y_n$ are i.i.d. uniformly bounded centered random matrices, and let $X = n^{-\frac{1}{2}}\sum_{i=1}^n Y_i$. By the central limit theorem, the distribution of X becomes Gaussian as $n \to \infty$. In this case $\sigma_1, \sigma_2, \sigma_*, v$ do not depend on n, but $L, R = O(n^{-\frac{1}{2}})$. Therefore, if we let $n \to \infty$, Theorem 2.12 reproduces the Gaussian bounds of Corollary 2.2 up to a universal constant (cf. Lemma 2.5).

3. Examples

The aim of this section is to illustrate our main results in concrete examples. In Section 3.1 we consider Gaussian random matrices with independent entries, while Section 3.2 discusses some simple examples of random matrix models with dependent entries. Section 3.3 is concerned with Gaussian sample covariance matrices, whose samples may be neither independent nor identically distributed.

3.1. **Independent entries.** In this section, we consider the case of real symmetric Gaussian random matrices with independent entries (nonsymmetric or complex matrices may be considered analogously, but we restrict attention to the real symmetric case for simplicity). More precisely, let X be the $d \times d$ symmetric random matrix with entries $X_{ij} = b_{ij}g_{ij}$, where $\{g_{ij} : i \ge j\}$ are i.i.d. standard real Gaussian random variables and $\{b_{ij} : i \geq j\}$ are given nonnegative scalars. We let $b_{ji} := b_{ij}$ and $g_{ji} := g_{ij}$. This model may be expressed in the form (2.1) as

$$X = \sum_{i \geq j} g_{ij} b_{ij} E_{ij}, \tag{3.1}$$

where $E_{ii} := e_i e_i^*$ and $E_{ij} := (e_i e_j^* + e_j e_i^*)$ for $i > j$. Here and in the sequel, $e_1, \ldots, e_d$ denotes the coordinate basis of $\mathbb{R}^d$.

The independent entry setting is the only general model of nonhomogeneous random matrices for which satisfactory norm bounds were obtained prior to this work [5, 38, 22]. In particular, it was proved in [5, Theorem 1.1] that

$$\mathbf{E}\|X\| \leq (2+\varepsilon) \max_i \sqrt{\sum_j b_{ij}^2} + \frac{C}{\sqrt{\varepsilon}} \max_{ij} b_{ij} \sqrt{\log d} \tag{3.2}$$

for any $0 < \varepsilon < 1$, where C is a universal constant. The constant 2 in the leading term is optimal, as $\mathbf{E}\|X\| = 2 + o(1)$ as $d \to \infty$ when X is a standard Wigner matrix, that is, when $b_{ij} = \frac{1}{\sqrt{d}}$ for all i, j. Moreover, (3.2) is nearly sharp in the sense that the inequality can be reversed up to a universal constant under mild assumptions [5, §3.5] (a completely sharp dimension-free bound, but without the optimal constant in the leading term, was proved in [22]).

Nonetheless, even in the special case of independent entries, the general results of this paper can yield a significant improvement over (3.2).

Lemma 3.1. *For the model* (3.1)*, we have*

$$\sigma(X) = \max_i \sqrt{\sum_j b_{ij}^2}, \qquad \max_{ij} b_{ij} \leq \sigma_*(X) \leq v(X) \leq \sqrt{2} \max_{ij} b_{ij}.$$

In particular,

$$\mathbf{E}\|X\| \leq (1+\varepsilon)\|X_{\mathrm{free}}\| + \frac{C}{\varepsilon} \max_{ij} b_{ij} \, (\log d)^{\frac{3}{2}} \tag{3.3}$$

for any $\varepsilon > 0$, where C is a universal constant.

Proof. The expression for $\sigma(X)^2 = \|\mathbf{E}X^2\|$ follows readily as

$$\mathbf{E}X^2 = \sum_i e_i e_i^* \sum_j b_{ij}^2 \tag{3.4}$$

is a diagonal matrix. Moreover, that $v(X)^2 \geq \sigma_*(X)^2 \geq \max_{ij} \mathbf{E}[|X_{ij}|^2] = \max_{ij} b_{ij}^2$ follows immediately from the definitions in Section 2.1.

On the other hand, as the pairs of entries (X_{ij}, X_{ji}) are independent for distinct indices $i \geq j$, we have $\mathrm{Cov}(X) = \bigoplus_{i \geq j} C_{ij}$ where C_{ij} is the covariance matrix of (X_{ij}, X_{ji}). Thus $v(X)^2 = \|\mathrm{Cov}(X)\| = \max_{i \geq j} \|C_{ij}\| \leq 2 \max_{ij} b_{ij}^2$.

To conclude, it remains to invoke Corollary 2.2 and to note that $c\tilde{v}(X)(\log d)^{\frac{3}{4}} \leq \varepsilon\|X_{\mathrm{free}}\| + \frac{c^2}{4\varepsilon} v(X)(\log d)^{\frac{3}{2}}$ for any $c, \varepsilon > 0$ by Young's inequality and Lemma 2.5. □

While the second term of (3.3) has a slightly suboptimal power on the logarithm as compared to (3.2), this term is already negligible when

$$\max_{ij} b_{ij}^2 = o\bigg((\log d)^{-3} \max_i \sum_j b_{ij}^2 \bigg). \tag{3.5}$$

As soon as this is the case, the bound (3.3) improves on (3.2) in that the leading term $2\sigma(X)$ is replaced by the sharp free probability quantity $\|X_{\mathrm{free}}\|$. We always have $\|X_{\mathrm{free}}\| \le 2\sigma(X)$ by Lemma 2.5, but this inequality often turns out to be strict in nonhomogeneous situations. To understand this phenomenon better, it is instructive to compute $\|X_{\mathrm{free}}\|$ in the present setting.

Lemma 3.2. *For the model* (3.1)*, we have*

$$\|X_{\mathrm{free}}\| = \inf_{x\in\mathbb{R}^d_{++}} \max_i \left\{\frac{1}{x_i} + \sum_j b_{ij}^2 x_j\right\} = 2 \sup_{w\in\Delta^{d-1}} \sum_i \sqrt{w_i \sum_j b_{ij}^2 w_j},$$

where $\mathbb{R}^d_{++}$ is the positive orthant and Δ^{d-1} is the standard simplex in $\mathbb{R}^d$. We always have $\|X_{\mathrm{free}}\| \le 2\sigma(X)$. If $B = (b_{ij}^2)$ is an irreducible nonnegative matrix, then equality $\|X_{\mathrm{free}}\| = 2\sigma(X)$ holds if and only if $\max_i \sum_j b_{ij}^2 = \min_i \sum_j b_{ij}^2$.

Remark 3.3. The irreducibility assumption entails no loss of generality. In the general case, we may write $B = \bigoplus_i B_i$ in terms of its irreducible components B_i, and $X_{\mathrm{free}} = \bigoplus_i X_{\mathrm{free},i}$ decomposes accordingly. As $\|X_{\mathrm{free}}\| = \max_i \|X_{\mathrm{free},i}\|$, the characterization of when $\|X_{\mathrm{free}}\| = 2\sigma(X)$ reduces to the irreducible case.

Proof of Lemma 3.2. Define

$$f(Z) := Z^{-1} + \sum_{i\ge j} b_{ij}^2 E_{ij} Z E_{ij}.$$

Fix any $Z > 0$ so that $f(Z)$ is a multiple of the identity. Then $f(Z) = \mathrm{diag}(f(Z))$, where $\mathrm{diag}(M)_{ij} := M_{ii}\delta_{ij}$. Using that $(Z^{-1})_{ii} \ge (Z_{ii})^{-1}$ (as $\|Z^{\frac12}e_i\|\|Z^{-\frac12}e_i\| \ge 1$), it follows readily that $f(Z) \ge f(\mathrm{diag}(Z))$. Thus Lemma 2.4 implies

$$\|X_{\mathrm{free}}\| = \inf_{Z>0} \lambda_{\max}(f(\mathrm{diag}(Z))) = \inf_{x\in\mathbb{R}^d_{++}} \max_i \left\{\frac{1}{x_i} + \sum_j b_{ij}^2 x_j\right\}.$$

We can further compute

$$\|X_{\mathrm{free}}\| = \inf_{x\in\mathbb{R}^d_{++}} \sup_{w\in\Delta^{d-1}} \sum_i w_i \left\{\frac{1}{x_i} + \sum_j b_{ij}^2 x_j\right\} = 2 \sup_{w\in\Delta^{d-1}} \sum_i \sqrt{w_i \sum_j b_{ij}^2 w_j},$$

where we used the Sion minimax theorem to exchange the infimum and supremum.

If we apply Cauchy-Schwarz to the rightmost expression for $\|X_{\mathrm{free}}\|$, we obtain $\|X_{\mathrm{free}}\| \le 2\max_i [\sum_j b_{ij}^2]^{\frac12} = 2\sigma(X)$ directly. Therefore, when $\|X_{\mathrm{free}}\| = 2\sigma(X)$, the maximizing vector $w \in \Delta^{d-1}$ must yield equality in Cauchy-Schwarz. The latter implies there exists $\rho \ge 0$ such that $Bw = \rho w$ and $\|X_{\mathrm{free}}\| = 2\sqrt{\rho}$. In particular, if B is irreducible, then $\rho = \rho(B)$ is the largest eigenvalue of B by the Perron-Frobenius theorem [16, p. 53]. It remains to recall that the inequality $\rho(B) \le \max_i \sum_j b_{ij}^2$ is strict unless $\max_i \sum_j b_{ij}^2 = \min_i \sum_j b_{ij}^2$, cf. [16, p. 63]. □

In other words, under the mild assumption (3.5), the constant 2 in (3.2) is suboptimal and the results of the present paper yield strictly better bounds on $\mathbf{E}\|X\|$ as soon as $\sum_j b_{ij}^2 \ne \sum_j b_{kj}^2$ for some i, k (and X does not decompose as a block-diagonal matrix). In such cases, Lemma 3.2 can be used to explicitly compute or estimate $\|X_{\mathrm{free}}\|$. The latter quantity has also been studied by completely different methods in [13], to which we refer for complementary results.

Even when $\max_i \sum_j b_{ij}^2 = \min_i \sum_j b_{ij}^2$, however, our main results yield far stronger conclusions than just a bound on the spectral norm. Indeed, by (3.4), this corresponds precisely to the case where $\mathbf{E}[X^2] = \sigma(X)^2\mathbf{1}$; thus any independent family of such matrices is strongly asymptotically free by Theorem 2.10.

Corollary 3.4. *Let $s_1, \ldots, s_m$ be a free semicircular family. For each $N \geq 1$, let $H_1^N, \ldots, H_m^N$ be independent random matrices of dimension $d = d(N) \geq N$ of the form (3.1), such that the variance pattern (b_{ij}^2) of H_k^N satisfies*

$$\max_i \sum_j b_{ij}^2 = \min_i \sum_j b_{ij}^2 = 1, \qquad \max_{ij} b_{ij}^2 = o((\log d)^{-3})$$

for every k, N. Then

$$\lim_{N\to\infty} \|p(H_1^N, \ldots, H_m^N)\| = \|p(s_1, \ldots, s_m)\| \quad a.s.,$$
$$\lim_{N\to\infty} \operatorname{tr} p(H_1^N, \ldots, H_m^N) = \tau(p(s_1, \ldots, s_m)) \quad a.s.$$

for every noncommutative polynomial p.

Corollary 3.4 provides a large class of new examples of strongly asymptotically free random matrices. Let us highlight a particularly interesting case.

Example 3.5 (Sparse Wigner matrices). Let $\mathrm{G} = ([d], E)$ be a k-regular graph with d vertices. A G-*sparse Wigner matrix* is a $d \times d$ real symmetric random matrix X such that $X_{ij} = k^{-\frac{1}{2}} g_{ij} 1_{\{i,j\}\in E}$ for $i \geq j$, where $\{g_{ij} : i \geq j\}$ are i.i.d. standard Gaussians. Note that X has only kd nonzero entries.

Now consider any sequence of k_N-regular graphs G_N with d_N vertices, and let $H_1^N, \ldots, H_m^N$ be independent G_N-sparse Wigner matrices. Then Corollary 3.4 shows that $H_1^N, \ldots, H_m^N$ are strongly asymptotically free as soon as $k_N \gg (\log d_N)^3$.

This example is striking for at least two reasons. First, all but a vanishing fraction of the entries of the matrices H_i^N are zero (for example, $d \log^4 d$ nonzero entries already suffice), so that strong asymptotic freeness is achieved here with far less randomness than is present in standard Wigner matrices. Second, no assumption whatsoever made on the graphs G_N except their regularity; in particular, the distributions of H_i^N need not possess any special symmetries. Let us note that even weak asymptotic freeness was previously known in the present setting only under very strong restrictions on the variance pattern, cf. [31, 3].

Beyond norm bounds and asymptotic freeness, applying Theorems 2.1 or 2.8 to the independent entry model (3.1) provides detailed information on the spectrum of X for arbitrary variance patterns b_{ij}^2 satisfying the mild assumption (3.5). In the interest of brevity we do not spell out these conclusions further.

3.2. **Dependent entries.** The aim of this section is to discuss some simple examples of random matrices with dependent entries. Unlike the independent entry model of the previous section, the only general nonasymptotic bound that was previously available in the dependent setting is the noncommutative Khintchine inequality (1.2) and analogous matrix concentration inequalities.

The following examples illustrate that, in many cases, our results are able to remove the dimensional factor in (1.2) under mild assumptions. To this end, note that for any random matrix X with centered jointly Gaussian entries, we have $\mathbf{E}\|X\| \gtrsim \sigma(X)$ by (1.2) and Remark 2.6. On the other hand, Corollary 2.2 and Lemma 2.5 imply that $\mathbf{E}\|X\| \lesssim \sigma(X)$ as soon as $v(X)(\log d)^{\frac{3}{2}} \lesssim \sigma(X)$. We aim to understand when the latter condition holds in concrete examples.

3.2.1. *Patterned random matrices.* Our first example is a model where independent Gaussians are placed in a matrix according to a given pattern. More precisely, let $g_1, \ldots, g_n$ be i.i.d. standard real Gaussian variables and let $S_1, \ldots, S_n$ be a partition of $[d] \times [d]$. We define X such that $X_{jk} = d^{-\frac{1}{2}} g_i$ for $(j,k) \in S_i$; thus

$$X = \sum_{i=1}^{n} g_i A_i, \qquad (A_i)_{jk} = \frac{1_{(j,k)\in S_i}}{\sqrt{d}}. \tag{3.6}$$

Many classical patterned random matrix models, such as random Toeplitz or Hankel matrices, are special cases of this model; cf. [7].

Lemma 3.6. *For the model (3.6), we have* $\mathbf{E}\|X\| \asymp \sigma(X)$ *when* $\max_i |S_i| \lesssim \frac{d}{(\log d)^3}$.

Proof. As $S_1, \ldots, S_n$ partition $[d] \times [d]$, we have

$$\sigma(X)^2 \geq \mathrm{tr}\left(\sum_i A_i^* A_i\right) = \frac{1}{d^2} \sum_i |S_i| = 1.$$

On the other hand, as $(X_{kl})_{(k,l)\in S_i}$ are independent for distinct i, we have $\mathrm{Cov}(X) = \bigoplus_i C_i$ where C_i is the covariance matrix of $(X_{kl})_{(k,l)\in S_i}$. Therefore

$$v(X)^2 = \|\mathrm{Cov}(X)\| = \max_i \|C_i\| = \max_i \frac{|S_i|}{d}.$$

The assumption now immediately implies $v(X)(\log d)^{\frac{3}{2}} \lesssim \sigma(X)$. $\square$

Lemma 3.6 shows that when $\max_i |S_i| \lesssim \frac{d}{(\log d)^3}$, the dimensional factor in the noncommutative Khintchine inequality (1.2) is unnecessary. On the other hand, Gaussian Toeplitz matrices provide an example with $\max_i |S_i| = d$ for which the dimensional factor in the noncommutative Khintchine inequality is necessary: in this case $\sigma(X) = 1$ and $\mathbf{E}\|X\| \asymp \sqrt{\log d}$ [35, §4.4]. Thus Lemma 3.6 is nearly the best one can hope for. This kind of "phase transition" between regimes where the noncommutative Khintchine inequality is and is not accurate is a common feature that will be observed in several other examples.

For a general choice of pattern $S_1, \ldots, S_n$, the parameter $\sigma(X)$ may be difficult to compute explicitly. However, for special choices of patterns we can obtain much stronger information. The following simple example provides a model where strong asymptotic freeness arises for matrices that contain many dependent entries.

Example 3.7 (Special patterned matrices). Suppose $S_1, \ldots, S_n$ satisfy the following:
1. Each S_i is symmetric (that is, $(k,l) \in S_i \Leftrightarrow (l,k) \in S_i$).
2. Each S_i has at most one entry in each row of $[d] \times [d]$.
3. $\max_i |S_i| \leq \frac{d}{(\log d)^4}$.

The first assumption implies that each A_i is a symmetric matrix. The second assumption implies that A_i^2 is a diagonal matrix; moreover,

$$(\mathbf{E}[X^2])_{kk} = \sum_i (A_i^2)_{kk} = \frac{1}{d} \sum_i 1_{S_i \text{ has an entry in row } k} = 1$$

for all k as $S_1, \ldots, S_n$ partition $[d] \times [d]$, so that $\mathbf{E}[X^2] = \mathbf{1}$. The third assumption implies that $v(X) \leq (\log d)^{-2}$. Matrices of this kind therefore satisfy the assumptions of Theorem 2.10. Thus if $H_1^d, \ldots, H_m^d$ are independent matrices satisfying the above assumptions, then they are strongly asymptotically free as $d \to \infty$.

3.2.2. *Independent columns.* Our second example is the model where the columns $X_1, \ldots, X_d$ of the random matrix X are independent centered Gaussian vectors with arbitrary covariance matrices $\Sigma_1, \ldots, \Sigma_d$. In this situation, all the relevant matrix parameters can be easily computed in explicit form.

Lemma 3.8. *For the independent columns model, we have*

$$\|\mathbf{E}[XX^*]\| = \left\| \sum_{i=1}^d \Sigma_i \right\|, \qquad \|\mathbf{E}[X^*X]\| = \max_i \mathrm{Tr}[\Sigma_i], \qquad v(X)^2 = \max_i \|\Sigma_i\|.$$

In particular,

$$\mathbf{E}\|X\| \leq (1+\varepsilon)\left\{ \left\| \sum_{i=1}^d \Sigma_i \right\|^{\frac{1}{2}} + \max_i \mathrm{Tr}[\Sigma_i]^{\frac{1}{2}} \right\} + \frac{C}{\varepsilon} \max_i \|\Sigma_i\|^{\frac{1}{2}} (\log d)^{\frac{3}{2}}$$

for any $\varepsilon > 0$, where C is a universal constant.

Proof. It follows readily from the definition of X that $\mathbf{E}[XX^*] = \sum_i \Sigma_i$, $\mathbf{E}[X^*X] = \sum_i \mathrm{Tr}[\Sigma_i]\, e_i e_i^*$, and $\mathrm{Cov}(X) = \bigoplus_i \Sigma_i$, which yields the first equation display. It remains to invoke Corollary 2.2, Lemma 2.5, and Young's inequality. □

Lemma 3.8 shows that we have $\mathbf{E}\|X\| \asymp \sigma(X)$ in the independent column model as soon as the last term in the norm bound is dominated by either of the first two terms. For example, this is the case if each Σ_i has sufficently large effective rank

$$\mathrm{rk}(\Sigma_i) := \frac{\mathrm{Tr}[\Sigma_i]}{\|\Sigma_i\|} \gtrsim (\log d)^3.$$

Conversely, when the effective rank is too small the dimensional factor in the non-commutative Khintchine inequality may be necessary: for example, in the special case $\Sigma_i = e_i e_i^*$ where X is a diagonal matrix with i.i.d. standard Gaussians on the diagonal, it is readily seen that $\sigma(X) = 1$ and $\mathbf{E}\|X\| \asymp \sqrt{\log d}$.

On the other hand, we may have $\mathbf{E}\|X\| \asymp \sigma(X)$ regardless of the effective rank when the first term in the norm bound dominates. For example, when X has i.i.d. columns, that is, when $\Sigma_1 = \cdots = \Sigma_d = \Sigma$, Lemma 3.8 implies

$$\mathbf{E}\|X\| \asymp \sqrt{d\|\Sigma\|} + \sqrt{\mathrm{Tr}\,\Sigma}.$$

This special case is well known, see, e.g., [39, Lemma 5.4].

3.2.3. *Independent blocks.* Our third example is the model

$$X = \begin{bmatrix} X^{1,1} & \cdots & X^{1,m} \\ \vdots & \ddots & \vdots \\ X^{m,1} & \cdots & X^{m,m} \end{bmatrix} \tag{3.7}$$

where $X^{i,j}$ are independent $r \times r$ random matrices.

Lemma 3.9. *Consider the model (3.7) where $X^{i,j}$ are independent centered Gaussian random matrices. Then we have*

$$\mathbf{E}\|X\| \leq (1+\varepsilon)\left\{ \max_i \left\| \sum_j \mathbf{E} X^{i,j}(X^{i,j})^* \right\|^{\frac{1}{2}} + \max_j \left\| \sum_i \mathbf{E}(X^{i,j})^* X^{i,j} \right\|^{\frac{1}{2}} \right\} + \frac{C}{\varepsilon} \max_{ij} v(X^{i,j})\, (\log rm)^{\frac{3}{2}}$$

for any $\varepsilon > 0$, where C is a universal constant.

Proof. A simple computation shows that $\|\mathbf{E}XX^*\| = \max_i \|\sum_j X^{i,j}(X^{i,j})^*\|$ and $\|\mathbf{E}X^*X\| = \max_j \|\sum_i (X^{i,j})^* X^{i,j}\|$. Moreover, as the blocks $X^{i,j}$ are independent, $\mathrm{Cov}(X) = \bigoplus_{i,j} \mathrm{Cov}(X^{i,j})$ and thus $v(X)^2 = \|\mathrm{Cov}(X)\| = \max_{i,j} v(X^{i,j})^2$. It remains to invoke Corollary 2.2, Lemma 2.5, and Young's inequality. □

The independent block model (3.7) may be viewed as intermediate between the independent entry model (3.1) and fully dependent random matrices. As a particularly simple example, consider the case where $X^{i,j}$ are all i.i.d. copies of the same centered Gaussian random matrix Z. Then Lemma 3.9 yields

$$\mathbf{E}\|X\| \lesssim \sqrt{m}\,\sigma(Z) + v(Z)\,(\log rm)^{\frac{3}{2}},$$

so that $\mathbf{E}\|X\| \asymp \sigma(X)$ as soon as $\sigma(Z)^2 \gtrsim \frac{\log(rm)^3}{m} v(Z)^2$. On the other hand, the case $m = 1$ encodes any centered Gaussian matrix, for which the dimensional factor of the noncommutative Khintchine inequality cannot be removed.

When the blocks $X^{i,j}$ are arbitrary (non-Gaussian) centered random matrices, we can still obtain nontrivial information from the matrix concentration inequality of Theorem 2.12. Let us illustrate this by means of a simple example. By a slight abuse of notation, we will still write $\sigma(X)^2 = \|\mathbf{E}[X^*X]\| \vee \|\mathbf{E}[XX^*]\|$, even though the matrix X in the following example is non-Gaussian.

Example 3.10 (Non-Gaussian i.i.d. blocks). Consider the model (3.7) where $X^{i,j}$ are i.i.d. copies of a centered $r \times r$ random matrix Z, and assume for simplicity that $\|Z\|_{\mathrm{HS}} = c$ a.s. (the constant $c > 0$ may depend on m, r). Then $\sigma(X)^2 = m\sigma(Z)^2$ and $\|\mathrm{Cov}(X)\| = \|\mathrm{Cov}(Z)\|$ as in the proof of Lemma 3.9. Moreover,

$$\|\mathrm{Cov}(X)\| \le \mathrm{Tr}[\mathrm{Cov}(Z)] = c^2 = \mathrm{Tr}[\mathbf{E}ZZ^*] \le r\sigma(Z)^2.$$

Thus Theorem 2.12 yields

$$\mathbf{E}\|X\| \lesssim \left(1 + (\log rm)^2 \sqrt{\frac{r}{m}}\right) \sigma(X).$$

In particular, $\mathbf{E}\|X\| \lesssim \sigma(X)$ as long as $r \lesssim \frac{m}{(\log m)^4}$.

On the other hand, when $r = m$ the conclusion may fail. For example, suppose the random matrix Z is uniformly distributed on the set $\{\pm e_1 e_1^*, \ldots, \pm e_r e_r^*\}$. Then $\sigma(X) = 1$, while $\mathbf{E}\|X\| \ge \mathbf{E}\max_i \|Xe_i\| \gtrsim (\frac{\log m}{\log\log m})^{\frac{1}{2}}$ [15, Theorem 3.4]. This example illustrates that a dimensional factor may be necessary in the matrix Bernstein inequality (2.2) when $r \gtrsim m$, while Theorem 2.12 shows that the dimensional factor can be eliminated when $r \lesssim \frac{m}{(\log m)^4}$.

3.2.4. *Gaussian on a subspace.* The examples discussed so far all feature a form of "structured independence", where certain subsets of entries are assumed to be independent. This is by no means necessary for the validity of our bounds. Our fourth example illustrates a simple situation that lacks any independence.

A matrix with i.i.d. real Gaussian entries may be viewed equivalently as the model defined by the isotropic Gaussian distribution on $\mathrm{M}_d(\mathbb{R})$. This model may generalized as follows. Let $\mathcal{M} \subseteq \mathrm{M}_d(\mathbb{R})$ be any linear subspace of dimension $\dim \mathcal{M} = k$ of the space of $d \times d$ real matrices, and let X be the random matrix defined by the isotropic Gaussian distribution on $\mathcal{M}$. Equivalently,

$$X = \sum_{i=1}^k g_i A_i$$

where $A_1,\ldots,A_k$ is any orthonormal basis of $\mathcal{M}$ (that is, $\mathrm{Tr}[A_i^*A_j]=\delta_{ij}$) and $g_1,\ldots,g_k$ are i.i.d. real standard Gaussian variables. Note that this model has fully dependent entries when $\mathcal{M}$ is in general position.

Lemma 3.11. *When X is an isotropic real Gaussian matrix on a linear subspace $\mathcal{M}\subseteq \mathrm{M}_d(\mathbb{R})$, we have $\mathbf{E}\|X\|\asymp\sigma(X)$ as soon as $\dim\mathcal{M}\gtrsim d\log^3 d$.*

Proof. Let $\dim\mathcal{M}=k$. Then $\sigma(X)^2\ge \mathrm{tr}[\sum_i A_i^*A_i]=\frac{k}{d}$. On the other hand, note that $\mathrm{Cov}(X)=\sum_{i=1}^n \iota(A_i)\iota(A_i)^*$, where $\iota:\mathrm{M}_d(\mathbb{R})\to\mathbb{R}^{d^2}$ maps a matrix to its vector of entries. But here $\iota(A_i)$ were assumed to be orthonormal, so $\mathrm{Cov}(X)$ is a projection matrix. Thus $v(X)^2=\|\mathrm{Cov}(X)\|=1$. As explained at the beginning of Section 3.2, We therefore have $\mathbf{E}\|X\|\asymp\sigma(X)$ as soon as $(\log d)^3\lesssim\frac{k}{d}$. □

When $\mathcal{M}=\mathrm{span}\{e_ie_j^*:|i-j|\le r\}$, we have $\dim\mathcal{M}\asymp(r+1)d$, $\sigma(X)\asymp\sqrt{r+1}$, and $\mathbf{E}\|X\|\ge\mathbf{E}\max_{ij}|X_{ij}|\gtrsim\sqrt{\log d}$. Thus the conclusion of Lemma 3.11 may fail when $\dim\mathcal{M}\ll d\log d$. While this particular example is rather special (as X has independent entries), the beauty of Lemma 3.11 is that it applies to *any* $\mathcal{M}$.

3.3. **Generalized sample covariance matrices.** Let X be any $d\times m$ random matrix with centered jointly Gaussian entries. We will refer to XX^* as a generalized sample covariance matrix. Indeed, as $\frac{1}{m}XX^*=\frac{1}{m}\sum_{i=1}^m X_iX_i^*$ in terms of the columns $X_1,\ldots,X_m$ of X, we see that $\frac{1}{m}XX^*$ is a sample covariance matrix in the special case that the data $X_1,\ldots,X_m$ are i.i.d. (see, e.g., [21]). In the general setting, one may still think of $\frac{1}{m}XX^*$ as a sample covariance matrix, but where the samples need not be independent or identically distributed.

The main question of interest in this setting is to estimate the deviation of the sample covariance matrix from the actual covariance matrix $\|XX^*-\mathbf{E}XX^*\|$. We presently show that an estimate of this kind can be derived from Theorem 2.1 using a simple variant of the linearization trick that is used in Theorem 2.10. While linearization generally yields asymptotic results for any polynomial, the present example illustrates that nonasymptotic bounds can be derived for specific polynomials by a careful analysis of the linearization argument. Alternatively, the interpolation method used in the proofs of our main results can be adapted directly to yield quantitative bounds for polynomials (we do not pursue this here).

Theorem 3.12. *Let $A_1,\ldots,A_n$ be arbitrary $d\times m$ matrices with complex entries, and define X and X_{free} as in (2.1) with $A_0=0$. Then we have*

$$\mathbf{E}\|XX^*-\mathbf{E}XX^*\|\le\|X_{\mathrm{free}}X_{\mathrm{free}}^*-\mathbf{E}XX^*\otimes\mathbf{1}\|\\ +C\{\sigma(X)\tilde{v}(X)(\log dm)^{\frac34}+\tilde{v}(X)^2(\log dm)^{\frac32}\},$$

where C is a universal constant.

The proof of Theorem 3.12 will be given at the end of this section. To clarify its meaning, it is instructive to note that $\mathbf{E}XX^*=(\mathrm{id}\otimes\tau)[X_{\mathrm{free}}X_{\mathrm{free}}^*]$; therefore, $\|X_{\mathrm{free}}X_{\mathrm{free}}^*-\mathbf{E}XX^*\otimes\mathbf{1}\|$ is precisely the free analogue of $\|XX^*-\mathbf{E}XX^*\|$.

To apply Theorem 3.12 in concrete situations, we must be able to compute or bound its right-hand side. The following bound often suffices.

Proposition 3.13. *In the setting of Theorem 3.12, we have*

$$\|X_{\mathrm{free}}X_{\mathrm{free}}^*-\mathbf{E}XX^*\otimes\mathbf{1}\|\le 2\|\mathbf{E}XX^*\|^{\frac12}\|\mathbf{E}X^*X\|^{\frac12}+\|\mathbf{E}X^*X\|.$$

Proof. We use the standard construction of a free semicircular family on Fock space, cf. [27, pp. 102–108] or [28, §9.9] (this construction will not be used elsewhere in this paper). Let $\mathcal{F}(\mathbb{C}^n) := \bigoplus_{k=0}^{\infty}(\mathbb{C}^n)^{\otimes k}$ be the free Fock space over $\mathbb{C}^n$ with creation operator $l(h)(x_1 \otimes \cdots \otimes x_k) := h \otimes x_1 \otimes \cdots \otimes x_k$ for $h \in \mathbb{C}^n$. Then the self-adjoint operators $s_1, \ldots, s_n$ defined by $s_i = l(e_i) + l(e_i)^*$ form a free semicircular family with respect to the vacuum state on $B(\mathcal{F}(\mathbb{C}^n))$.

As we assumed $A_0 = 0$, we may represent $X_{\text{free}} = U + V$ with $U := \sum_i A_i \otimes l(e_i)$ and $V := \sum_i A_i \otimes l(e_i)^*$. The property $l(e_i)^* l(e_j) = \delta_{ij}\mathbf{1}$ (which is readily verified from the definition of $l(h)$) yields the identities

$$VV^* = \sum_i A_i A_i^* \otimes \mathbf{1} = \mathbf{E}XX^* \otimes \mathbf{1}, \qquad U^*U = \sum_i A_i^* A_i \otimes \mathbf{1} = \mathbf{E}X^*X \otimes \mathbf{1}.$$

We therefore obtain

$$\begin{aligned}\|X_{\text{free}}X_{\text{free}}^* - \mathbf{E}XX^* \otimes \mathbf{1}\| &= \|UV^* + VU^* + UU^*\| \le 2\|U\|\|V\| + \|U\|^2 \\ &= 2\|\mathbf{E}X^*X\|^{\frac{1}{2}}\|\mathbf{E}XX^*\|^{\frac{1}{2}} + \|\mathbf{E}X^*X\|,\end{aligned}$$

completing the proof. □

To illustrate these bounds, consider the case where the columns of X are i.i.d. centered Gaussian vectors with covariance Σ (so that $\frac{1}{m}XX^*$ is a classical sample covariance matrix). Then Theorem 3.12 and Proposition 3.13 yield

$$\begin{aligned}\mathbf{E}\|\tfrac{1}{m}XX^* - \Sigma\| \le \|\Sigma\|\left\{2\sqrt{\frac{\text{rk}(\Sigma)}{m}} + \frac{\text{rk}(\Sigma)}{m}\right\} + \\ C\|\Sigma\|\left\{\left(1 \vee \frac{\text{rk}(\Sigma)}{m}\right)^{\frac{3}{4}}\frac{(\log dm)^{\frac{3}{4}}}{m^{\frac{1}{4}}} + \left(1 \vee \frac{\text{rk}(\Sigma)}{m}\right)^{\frac{1}{2}}\frac{(\log dm)^{\frac{3}{2}}}{m^{\frac{1}{2}}}\right\}\end{aligned}$$

where $\text{rk}(\Sigma) := \text{Tr}[\Sigma]/\|\Sigma\|$, and we used Lemma 3.8 to compute $\sigma(X)$ and $v(X)$. The leading term in this bound dominates when $\text{rk}(\Sigma)$ is not too small. The latter restriction is not optimal: it was shown in [21] that when X has i.i.d. columns, $\mathbf{E}\|\frac{1}{m}XX^* - \Sigma\|$ always agrees with the leading term in the above inequality up to a universal constant. On the other hand, our general bounds apply to arbitrary nonhomogeneous random matrices X, which are out of reach of previous methods. (In the special case that X has independent Gaussian entries, bounds similar to those of the present section were obtained in [8].)

We now turn to the proof of Theorem 3.12. The key idea is the following lemma, which provides an explicit linearization of the polynomial $(X, X^*) \mapsto XX^* + A$.

Lemma 3.14. *Let $A_\varepsilon = (\|\mathbf{E}XX^*\| + 4\varepsilon^2)\mathbf{1} - \mathbf{E}XX^*$, and define*

$$\breve{X}_\varepsilon = \begin{bmatrix} 0 & 0 & X & A_\varepsilon^{\frac{1}{2}} \\ 0 & 0 & 0 & 0 \\ X^* & 0 & 0 & 0 \\ A_\varepsilon^{\frac{1}{2}} & 0 & 0 & 0 \end{bmatrix}, \qquad \breve{X}_{\text{free},\varepsilon} = \begin{bmatrix} 0 & 0 & X_{\text{free}} & A_\varepsilon^{\frac{1}{2}} \otimes \mathbf{1} \\ 0 & 0 & 0 & 0 \\ X_{\text{free}}^* & 0 & 0 & 0 \\ A_\varepsilon^{\frac{1}{2}} \otimes \mathbf{1} & 0 & 0 & 0 \end{bmatrix}.$$

Then we have

$$\begin{aligned}&\text{sp}(\breve{X}_\varepsilon) \subseteq \text{sp}(\breve{X}_{\text{free},\varepsilon}) + [-\varepsilon, \varepsilon] \qquad \Longrightarrow \\ &\begin{cases}\lambda_+(XX^* + A_\varepsilon)^{\frac{1}{2}} \le \lambda_+(X_{\text{free}}X_{\text{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} + \varepsilon, \\ \lambda_-(XX^* + A_\varepsilon)^{\frac{1}{2}} \ge \lambda_-(X_{\text{free}}X_{\text{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} - \varepsilon\end{cases}\end{aligned}$$

for any $\varepsilon > 0$, where $\lambda_+(Z) := \sup \mathrm{sp}(Z)$ and $\lambda_-(Z) := \inf \mathrm{sp}(Z)$.

Proof. By Remark 2.6, we have

$$\mathrm{sp}(\breve{X}_\varepsilon) \cup \{0\} = \mathrm{sp}((XX^* + A_\varepsilon)^{\frac{1}{2}}) \cup -\mathrm{sp}((XX^* + A_\varepsilon)^{\frac{1}{2}}) \cup \{0\},$$

and analogously for $\breve{X}_{\mathrm{free},\varepsilon}$. If $\mathrm{sp}(\breve{X}_\varepsilon) \subseteq \mathrm{sp}(\breve{X}_{\mathrm{free},\varepsilon}) + [-\varepsilon, \varepsilon]$, then clearly

$$\lambda_+(XX^* + A_\varepsilon)^{\frac{1}{2}} \le \lambda_+(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} + \varepsilon.$$

On the other hand, as $\breve{X}_{\mathrm{free},\varepsilon}$ can have a zero eigenvalue, it follows that either

$$\lambda_-(XX^* + A_\varepsilon)^{\frac{1}{2}} \ge \lambda_-(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} - \varepsilon$$

or $\lambda_-(XX^* + A_\varepsilon)^{\frac{1}{2}} \le \varepsilon$. But the latter is impossible, as $\lambda_-(XX^* + A_\varepsilon)^{\frac{1}{2}} \ge 2\varepsilon$. $\square$

We can now complete the proof of Theorem 3.12.

Proof of Theorem 3.12. We adopt throughout the proof the notation and conclusions Lemma 3.14. By Remark 2.6, we have $\sigma_*(\breve{X}_\varepsilon) = \sigma_*(X)$ and $\tilde{v}(\breve{X}_\varepsilon) \le 2^{\frac{1}{4}}\tilde{v}(X)$. We may therefore apply Theorem 2.1 to $\breve{X}_\varepsilon$ to obtain

$$\begin{aligned}\mathbf{P}\big[&\lambda_+(XX^* + A_{\varepsilon(t)})^{\frac{1}{2}} \le \lambda_+(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_{\varepsilon(t)} \otimes \mathbf{1})^{\frac{1}{2}} + \varepsilon(t),\\ &\lambda_-(XX^* + A_{\varepsilon(t)})^{\frac{1}{2}} \ge \lambda_-(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_{\varepsilon(t)} \otimes \mathbf{1})^{\frac{1}{2}} - \varepsilon(t)\big] \ge 1 - e^{-t^2}\end{aligned}$$

for all $t \ge 0$, where $\varepsilon(t) = c\{\tilde{v}(X)(\log dm)^{\frac{3}{4}} + \sigma_*(X)t\}$ for a universal constant c.

Now note that

$$\lambda_\pm(XX^* + A_\varepsilon) = \lambda_\pm(XX^* - \mathbf{E}XX^*) + \|\mathbf{E}XX^*\| + 4\varepsilon^2,$$

and analogously for X_{free}. Moreover, we have

$$\lambda_-(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1}) \le \lambda_+(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1}) \le 5\sigma(X)^2 + 4\varepsilon^2$$

by Lemma 2.5. Thus we obtain

$$\begin{aligned}&\lambda_+(XX^* + A_\varepsilon)^{\frac{1}{2}} \le \lambda_+(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} + \varepsilon \qquad \Longrightarrow\\ &\lambda_+(XX^* - \mathbf{E}XX^*) \le \lambda_+(X_{\mathrm{free}}X_{\mathrm{free}}^* - \mathbf{E}XX^* \otimes \mathbf{1}) + 2\varepsilon\sqrt{5\sigma(X)^2 + 4\varepsilon^2} + \varepsilon^2\end{aligned}$$

by squaring both sides of the first inequality and applying the previous two equation displays. Analogously, using $(y - \varepsilon)_+^2 \ge y^2 - 2\varepsilon y - \varepsilon^2$ for $y, \varepsilon \ge 0$ yields

$$\begin{aligned}&\lambda_-(XX^* + A_\varepsilon)^{\frac{1}{2}} \ge \lambda_-(X_{\mathrm{free}}X_{\mathrm{free}}^* + A_\varepsilon \otimes \mathbf{1})^{\frac{1}{2}} - \varepsilon \qquad \Longrightarrow\\ &\lambda_-(XX^* - \mathbf{E}XX^*) \ge \lambda_-(X_{\mathrm{free}}X_{\mathrm{free}}^* - \mathbf{E}XX^* \otimes \mathbf{1}) - 2\varepsilon\sqrt{5\sigma(X)^2 + 4\varepsilon^2} - \varepsilon^2.\end{aligned}$$

But as $\|Z\| = \max(\lambda_+(Z), -\lambda_-(Z))$, we have shown that

$$\mathbf{P}\big[\|XX^* - \mathbf{E}XX^*\| > \|X_{\mathrm{free}}X_{\mathrm{free}}^* - \mathbf{E}XX^* \otimes \mathbf{1}\| + 5\sigma(X)\varepsilon(t) + 5\varepsilon(t)^2\big] \le e^{-t^2}.$$

The conclusion follows by integrating this tail bound and using $\sigma_*(X) \le \tilde{v}(X)$. $\square$

4. Preliminaries

The aim of this section is to recall some mathematical background and to introduce a few basic estimates that will be used in the remainder of the paper.

4.1. **Free probability.** We begin by recalling some basic notions of free probability; the reader is referred to [27] for an introduction to this topic.

For our purposes, a *unital C^*-algebra* may be thought of concretely as an algebra $\mathcal{A}$ of bounded operators on a complex Hilbert space which is self-adjoint ($a \in \mathcal{A}$ implies $a^* \in \mathcal{A}$), is closed in the operator norm, and contains the identity $\mathbf{1} \in \mathcal{A}$. A *state* is a linear functional $\tau : \mathcal{A} \to \mathbb{C}$ that is positive $\tau(a^*a) \geq 0$ and unital $\tau(\mathbf{1}) = 1$. A state is called *faithful* if $\tau(a^*a) = 0$ implies $a = 0$.

Definition 4.1. A *C^*-probability space* is a pair $(\mathcal{A}, \tau)$, where $\mathcal{A}$ is a unital C^*-algebra and τ is a faithful state.

The simplest example of a C^*-probability space is $(\mathrm{M}_d(\mathbb{C}), \mathrm{tr})$. The introduction of general C^*-probability spaces enables us to extend computations involving matrices and traces to infinite-dimensional operators. The assumption that τ is faithful ensures that $\|a\| = \lim_{p\to\infty} \tau(|a|^p)^{\frac{1}{p}}$ [27, Proposition 3.17].

The basic infinite-dimensional object of interest in this paper is a free semicircular family. We will define this notion combinatorially as in [27, p. 128]. For any integer p, denote by $\mathrm{P}_2([p])$ the collection of all pairings of $[p] := \{1, \ldots, p\}$, that is, of partitions of $[p]$ each of whose blocks consists of exactly two elements. We denote by $\mathrm{NC}_2([p]) \subseteq \mathrm{P}_2([p])$ the collection of those pairings π that are *noncrossing*, i.e., that do not contain $\{i, j\}, \{k, l\} \in \pi$ so that $i < k < j < l$.

Definition 4.2. A family $s_1, \ldots s_n \in \mathcal{A}$ of self-adjoint elements in a C^*-probability space $(\mathcal{A}, \tau)$ is called a *free semicircular family* if

$$\tau(s_{k_1} \cdots s_{k_p}) = \sum_{\pi \in \mathrm{NC}_2([p])} \prod_{\{i,j\} \in \pi} \delta_{k_i k_j}$$

for every $p \geq 1$, $k_1, \ldots, k_p \in [n]$.

The elements s_i are "semicircular" in the sense that for $p \in \mathbb{N}$,

$$\tau(s_i^p) = |\mathrm{NC}_2([p])| = \int_{-2}^{2} x^p \cdot \tfrac{1}{2\pi} \sqrt{4 - x^2}\, dx$$

are the moments of the standard semicircle distribution, cf. [27, p. 123 and p. 29]. The latter is precisely the limiting spectral distribution of large Wigner matrices. In particular, note that $\|s_i\| = \lim_{p\to\infty} \tau(s_i^{2p})^{\frac{1}{2p}} = 2$.

More generally, the weak asymptotic freeness theorem of Voiculescu [40] states that a free semicircular family arises as the limiting object associated to independent Wigner matrices. A self-contained proof of this fact may be readily obtained as a special case of the argument in Section 7.1 below.

Theorem 4.3 (Voiculescu)**.** *Let $G_1^N, \ldots, G_n^N$ be independent standard Wigner matrices in the sense of Definition 1.1. Then we have*

$$\lim_{N\to\infty} \mathbf{E}[\mathrm{tr}(G_{k_1}^N \cdots G_{k_p}^N)] = \tau(s_{k_1} \cdots s_{k_p})$$

for every $p \geq 1$, $k_1, \ldots, k_p \in [n]$.

We now turn our attention to the basic random matrix model (2.1) of this paper. In the proofs of our main results, it will suffice to consider self-adjoint coefficient matrices $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ due to Remark 2.6. In addition to X and X_{free}

defined in (2.1), we also introduce the intermediate model

$$X^N := A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes G_i^N, \tag{4.1}$$

where $G_1^N, \ldots, G_n^N$ are independent standard Wigner matrices of dimension N. Theorem 4.3 enables us to compute the limiting spectral statistics of X^N.

Corollary 4.4. *Let $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$. Then*

$$\lim_{N\to\infty} \mathbf{E}[\mathrm{tr}\, f(X^N)] = (\mathrm{tr} \otimes \tau)(f(X_{\mathrm{free}}))$$

for any polynomial or bounded continuous function $f : \mathbb{R} \to \mathbb{C}$.

Proof. For the function $f(x) = x^p$ with $p \in \mathbb{N}$, we compute explicitly

$$\mathbf{E}\,\mathrm{tr}[(X^N)^p] = \sum_{i_1,\ldots,i_p=1}^n \mathrm{tr}(A_{i_1} \cdots A_{i_p}) \mathbf{E}[\mathrm{tr}\, G_{i_1} \cdots G_{i_p}] \xrightarrow{N\to\infty} (\mathrm{tr} \otimes \tau)(X_{\mathrm{free}}^p)$$

by Theorem 4.3. The conclusion extends to any polynomial f by linearity. For bounded continuous f, it remains to note that as $\|X_{\mathrm{free}}\| \le 2\sum_{i=0}^n \|A_i\| < \infty$, moment convergence implies weak convergence [27, p. 116]. □

We finally discuss a number of methods to compute or estimate the spectral statistics of X_{free}. First, we note that the moments of X_{free} are readily computed using Definition 4.2: for every $p \in \mathbb{N}$, we obtain

$$(\mathrm{tr} \otimes \tau)(X_{\mathrm{free}}^p) = \sum_{\pi \in \mathrm{NC}_2([p])} \sum_{(i_1,\ldots,i_p)\sim\pi} \mathrm{tr}(A_{i_1} \cdots A_{i_p}), \tag{4.2}$$

where $(i_1, \ldots, i_p) \sim \pi$ denotes that $i_k = i_l$ for every $\{k, l\} \in \pi$.

An explicit expression for the norm $\|X_{\mathrm{free}}\|$ was given in Lemma 2.4 above. This fundamental result was proved by Lehner [25, Corollary 1.5], where it is formulated only in the case that $A_0 \ge 0$ is positive semidefinite. However, the general formulation is readily derived from this special case.

Proof of Lemma 2.4. We first note that $t := \|X_{\mathrm{free}}\| \ge \|(\mathrm{id} \otimes \tau)(X_{\mathrm{free}})\| = \|A_0\|$. Thus $X_{\mathrm{free}} + t\mathbf{1} \ge 0$ and $A_0 + t\mathbf{1} \ge 0$. Applying [25, Corollary 1.5] yields

$$\|X_{\mathrm{free}} + t\mathbf{1}\| = \inf_{Z>0} \left\| Z^{-1} + A_0 + t\mathbf{1} + \sum_{i=1}^n A_i Z A_i \right\|,$$

where the infimum may be further restricted to Z for which the matrix in the norm on the right-hand side is a multiple of the identity. But as $X_{\mathrm{free}} + t\mathbf{1} \ge 0$, we have $\|X_{\mathrm{free}} + t\mathbf{1}\| = \lambda_{\max}(X_{\mathrm{free}}) + t$, and analogously for the norm on the right-hand side. It remains to use that $\|X_{\mathrm{free}}\| = \lambda_{\max}(X_{\mathrm{free}}) \vee -\lambda_{\max}(-X_{\mathrm{free}})$. □

Finally, the estimates on $\|X_{\mathrm{free}}\|$ in Lemma 2.5 were proved by Pisier [28, p. 208] in the case $A_0 = 0$ (the proof is very similar to that of Proposition 3.13 above). The extension to general A_0 follows immediately, however, using $\|A_0\| \le \|X_{\mathrm{free}}\| \le \|X_{\mathrm{free}} - A_0 \otimes \mathbf{1}\| + \|A_0\|$ (the first inequality was explained above in the proof of Lemma 2.4, and the second is the triangle inequality).

4.2. **Matrix parameters.** The aim of this section is to develop some basic properties of the parameters $\sigma(X), \sigma_*(X), v(X)$ defined in Section 2.1, and of the matrix alignment parameter $w(X)$ that was defined in Section 1.4.

4.2.1. *The matrix alignment parameter.* We will in fact need a somewhat more general parameter than $w(X)$ in our proofs, so we begin by defining the relevant notion. Let $A_0,\ldots,A_n,A_0',\ldots,A_n' \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$, and define the random matrices $X = A_0 + \sum_{i=1}^n g_i A_i$ and $X' = A_0' + \sum_{i=1}^n g_i A_i'$ as in (2.1). We define

$$w(X,X') := \sup_{U,V,W} \left\| \sum_{i,j=1}^n A_i U A_j' V A_i W A_j' \right\|^{\frac{1}{4}},$$

where the supremum is taken over all unitary matrices $U,V,W \in \mathrm{M}_d(\mathbb{C})$. Note that the definition of $w(X,X')$ does not involve A_0, A_0', and that $w(X,X') = w(X',X)$ (by taking the adjoint of inside the norm). Note also that we only defined $w(X,X')$ for self-adjoint coefficient matrices A_i, A_i'; the definition may be generalized to non-self-adjoint matrices, but this will not be needed in the sequel. In agreement with the notation of Section 1.4, we let $w(X) := w(X,X)$.

The matrix alignment parameter $w(X)$ was introduced by Tropp in [37] to quantify the contribution of crossings to the moments of X. A key idea of [37] is that upper bounds in terms of $w(X)$ may be obtained by complex interpolation. The following variant of this idea suffices for our purposes.

Lemma 4.5. *Let $Y^{(1)},\ldots,Y^{(4)}$ be arbitrary $d\times d$ complex random matrices, and let $p_1,\ldots,p_4 \geq 1$ satisfy $\sum_{k=1}^4 \frac{1}{p_k} = 1$. Then we have*

$$\left| \sum_{i,j=1}^n \mathbf{E}[\mathrm{tr}\, A_i Y^{(1)} A_j' Y^{(2)} A_i Y^{(3)} A_j' Y^{(4)}] \right| \leq w(X,X')^4 \prod_{k=1}^4 \mathbf{E}[\mathrm{tr}\, |Y^{(k)}|^{p_k}]^{\frac{1}{p_k}}.$$

Proof. We aim to show that $F(Y_1,\ldots,Y_4) := \sum_{i,j} \mathbf{E}[\mathrm{tr}\, A_i Y_1 A_j' Y_2 A_i Y_3 A_j' Y_4]$ satisfies $|F(Y_1,\ldots,Y_4)| \leq w(X,X')^4 \|Y_1\|_{p_1}\cdots\|Y_4\|_{p_4}$, where $\|Y\|_p := \mathbf{E}[\mathrm{tr}\,|Y|^p]^{\frac{1}{p}}$ denotes the $L_p(S_p)$-norm. Recall that the spaces $L_p(S_p)$ form a complex interpolation scale $L_r(S_r) = (L_p(S_p), L_q(S_q))_\theta$ with $\frac{1}{r} = \frac{1-\theta}{p} + \frac{\theta}{q}$ [29, §2]. By the classical complex interpolation theorem for multilinear maps [9, §10.1], it suffices to prove the conclusion in the case that $p_i = 1$ for some i. By cyclic permutation of the trace, we may assume $p_4 = 1$ and $p_1,p_2,p_3 = \infty$. But in this case

$$\sup_{\substack{\|Y_k\|_\infty \leq 1 \\ k=1,2,3}} \sup_{\|Y_4\|_1 \leq 1} |F(Y_1,\ldots,Y_4)| = \sup_{\substack{\|Y_k\| \leq 1 \\ k=1,2,3}} \left\| \sum_{i,j=1}^n A_i Y_1 A_j' Y_2 A_i Y_3 A_j' \right\| = w(X,X')^4$$

follows from the fact that every $Y \in \mathrm{M}_d(\mathbb{C})$ with $\|Y\| \leq 1$ is a convex combination of unitaries (by singular value decomposition and the fact that any vector $x \in \mathbb{R}^n$ with $\|x\|_\infty \leq 1$ is a convex combination of vectors in $\{-1,+1\}^d$). □

4.2.2. *Bounding the matrix alignment parameter.* The aim of this section is to prove the following bound on the matrix alignment parameter.

Proposition 4.6. *We have $w(X,X')^4 \leq v(X)\sigma(X)v(X')\sigma(X')$.*

To this end, we will require two simple observations.

Lemma 4.7. *In the proof of Proposition 4.6, there is no loss of generality in assuming that $\mathrm{Tr}[A_iA_j] = 0$ and $\mathrm{Tr}[A_i'A_j'] = 0$ for all $i \neq j$. In particular, this assumption implies $v(X) = \max_i \|A_i\|_{\mathrm{HS}}$ and $v(X') = \max_i \|A_i'\|_{\mathrm{HS}}$.*

Proof. It is evident from the definitions that the parameters $\sigma(X), v(X), w(X,X')$ only depend on the distributions of the random matrices X, X', and not on their representations in terms of A_i, A_i'. It therefore suffices to find random matrices Y, Y' that are equidistributed with X, X' and satisfy the desired properties.

To this end, note first that $\mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ is a real vector space of dimension d^2, endowed with the Hilbert-Schmidt inner product. Moreover, the distribution of X is a real Gaussian measure on this space. If we denote by $C_1, \ldots, C_{d^2} \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ the (unnormalized) orthogonal eigenvectors of the corresponding covariance matrix, it follows that X has the same distribution as $Y = A_0 + \sum_i g_i C_i$, and $\mathrm{Tr}[C_iC_j] = 0$ for $i \neq j$ by construction. Finally, note that $\mathrm{Cov}(Y) = \sum_i \iota(C_i)\iota(C_i)^*$, where $\iota : \mathrm{M}_d(\mathbb{C}) \to \mathbb{C}^{d^2}$ maps a matrix to its vector of entries. As the vectors $\iota(C_i)$ are orthogonal in $\mathbb{C}^{d^2}$, they are also eigenvectors of $\mathrm{Cov}(Y)$. It follows that $v(Y)^2 = \|\mathrm{Cov}(Y)\| = \max_i \|C_i\|_{\mathrm{HS}}^2$. The analogous construction applies to X'. □

Lemma 4.8. *Let $B_1, \ldots, B_{d^2} \in \mathrm{M}_d(\mathbb{C})$ satisfy $\mathrm{Tr}[B_i^* B_j] = \delta_{ij}$ for all $1 \leq i \leq j \leq n$. Then we have $\sum_{i=1}^{d^2} B_i^* Y B_i = \mathrm{Tr}[Y]\mathbf{1}$ for every $Y \in \mathrm{M}_d(\mathbb{C})$.*

Proof. Note that $\sum_{i=1}^{d^2} B_i^* Y B_i = \mathbf{E} H^* Y H$, where $H = \sum_{i=1}^{d^2} h_i B_i$ and $h_1, \ldots, h_{d^2}$ are i.i.d. standard complex Gaussians. Thus by unitary invariance of the complex Gaussian distribution, we may replace $B_1, \ldots, B_{d^2}$ by any other orthonormal basis of $\mathrm{M}_d(\mathbb{C})$. It follows that $\sum_{i=1}^{d^2} B_i^* Y B_i = \sum_{k,l=1}^{d} e_k e_l^* Y e_l e_k^* = \mathrm{Tr}[Y]\mathbf{1}$. □

We now complete the proof of Proposition 4.6.

Proof of Proposition 4.6. By Lemma 4.7, we can assume that $\mathrm{Tr}[A_iA_j] = 0$ and $\mathrm{Tr}[A_i'A_j'] = 0$ for $i \neq j$. In particular, we may choose an orthonormal basis $B_1, \ldots, B_{d^2}$ of $\mathrm{M}_d(\mathbb{C})$ so that $A_i = \|A_i\|_{\mathrm{HS}} B_i$ for $i = 1, \ldots, n$.

Now note that we can estimate by Cauchy-Schwarz

$$\begin{aligned} w(X,X')^4 &= \sup_{U,V,W} \sup_{\|x\|,\|y\| \leq 1} \left| \sum_{i=1}^n \left\langle U^* A_i x, \sum_{j=1}^n A_j' V A_i W A_j' y \right\rangle \right| \\ &\leq \left(\sup_{\|x\| \leq 1} \sum_{i=1}^n \|A_i x\|^2 \right)^{\frac{1}{2}} \left(\sup_{V,W} \sup_{\|y\| \leq 1} \sum_{i=1}^n \left\| \sum_{j=1}^n A_j' V A_i W A_j' y \right\|^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Furthermore,

$$\begin{aligned} \sum_{i=1}^n \left\| \sum_{j=1}^n A_j' V A_i W A_j' y \right\|^2 &\leq \max_i \|A_i\|_{\mathrm{HS}}^2 \sum_{i=1}^{d^2} \left\| \sum_{j=1}^n A_j' V B_i W A_j' y \right\|^2 \\ &= \max_i \|A_i\|_{\mathrm{HS}}^2 \sum_{j,k=1}^n \langle y, A_j' A_k' y \rangle \, \mathrm{Tr}[A_j' A_k'] \\ &\leq \max_i \|A_i\|_{\mathrm{HS}}^2 \max_i \|A_i'\|_{\mathrm{HS}}^2 \sum_{j=1}^n \|A_j' y\|^2, \end{aligned}$$

where we used Lemma 4.8 in the equality and $\mathrm{Tr}[A_i'A_j'] = 0$ for $i \neq j$ in the second inequality. It remains to note that $\sup_{\|x\| \leq 1} \sum_{i=1}^n \|A_i x\|^2 = \sigma(X)^2$ and $\max_i \|A_i\|_{\mathrm{HS}} = v(X)$ by Lemma 4.7, and analogously for X'. □

4.2.3. *Self-adjoint dilation.* While we defined $w(X, X')$ only for self-adjoint X, X', we may extend the resulting inequalities to the general case by self-adjoint dilation as explained in Remark 2.6. For completeness, we presently provide proofs of the claims made in Remark 2.6. We first prove the following.

Lemma 4.9. *Let T be a bounded operator on a Hilbert space H, and denote by $\breve{T}$ the self-adjoint operator on $H \oplus H$ defined by*

$$\breve{T} = \begin{bmatrix} 0 & T \\ T^* & 0 \end{bmatrix}.$$

Then $\mathrm{sp}(\breve{T}) \cup \{0\} = \mathrm{sp}(|T|) \cup -\mathrm{sp}(|T|) \cup \{0\}$.

Proof. Let $T = V|T|$ be the polar decomposition of T, where V is a partial isometry with initial space $(\ker T)^\perp$ and final space $\mathrm{cl}(\mathrm{ran}\, T) = (\ker T^*)^\perp$. As $TT^* = V|T|^2V^* = VT^*TV^*$, it follows that $\mathrm{sp}(T^*T) \cup \{0\} = \mathrm{sp}(TT^*) \cup \{0\}$. Thus

$$\breve{T}^2 = \begin{bmatrix} TT^* & 0 \\ 0 & T^*T \end{bmatrix} \tag{4.3}$$

implies that $\mathrm{sp}(|\breve{T}|) \cup \{0\} = \mathrm{sp}(|T|) \cup \{0\}$. On the other hand, as

$$U^*\breve{T}U = -\breve{T}, \qquad U = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{bmatrix}$$

and U is unitary, we have $\mathrm{sp}(\breve{T}) = -\mathrm{sp}(\breve{T})$. The conclusion follows. □

We now verify that $\sigma(X), \sigma_*(X), v(X)$ are well behaved under dilation.

Lemma 4.10. *In the setting of Remark 2.6, we have*

$$\sigma(\breve{X}) = \sigma(X), \qquad \sigma_*(\breve{X}) = \sigma_*(X), \qquad v(X) \le v(\breve{X}) \le \sqrt{2}\, v(X).$$

Proof. We begin by noting that by (4.3)

$$\sigma(\breve{X})^2 = \|\mathbf{E}\breve{X}^2\| = \left\| \begin{bmatrix} \mathbf{E}XX^* & 0 \\ 0 & \mathbf{E}X^*X \end{bmatrix} \right\| = \sigma(X)^2.$$

Next, note that

$$\sigma_*(\breve{X})^2 = \sup_{\|v_1\|^2+\|v_2\|^2=1} \sup_{\|w_1\|^2+\|w_2\|^2=1} \mathbf{E}[|\langle v_1, Xw_2\rangle + \langle v_2, X^*w_1\rangle|^2].$$

Thus clearly $\sigma_*(\breve{X}) \ge \sigma_*(X)$, while by the triangle inequality

$$\sigma_*(\breve{X}) \le \sigma_*(X) \sup_{\|v_1\|^2+\|v_2\|^2=1} \sup_{\|w_1\|^2+\|w_2\|^2=1} (\|v_1\|\|w_2\| + \|v_2\|\|w_1\|) = \sigma_*(X).$$

Finally, note that

$$v(\breve{X})^2 = \sup_{\|M\|_{\mathrm{HS}}^2+\|N\|_{\mathrm{HS}}^2=1} \mathbf{E}[|\mathrm{Tr}[XM] + \mathrm{Tr}[X^*N]|^2],$$

so that $v(X) \le v(\breve{X}) \le \sqrt{2}\, v(X)$ follows in the same manner as for $\sigma_*(X)$. □

4.3. **Gaussian analysis.** We now recall some Gaussian tools that will be used in the sequel. The following is classical [32, Lemma 1.3.1].

Lemma 4.11 (Gaussian interpolation)**.** *Let Y and Z be independent centered Gaussian vectors in $\mathbb{R}^n$ with covariance matrices Σ^Y and Σ^Z, respectively. Let*

$$Y_t = \sqrt{t}\,Y + \sqrt{1-t}\,Z$$

for $t \in [0,1]$. Then we have

$$\frac{d}{dt}\mathbf{E}[f(Y_t)] = \frac{1}{2}\sum_{i,j=1}^{n}(\Sigma^Y_{ij} - \Sigma^Z_{ij})\,\mathbf{E}\bigg[\frac{\partial^2 f}{\partial x_i \partial x_j}(Y_t)\bigg]$$

for any smooth $f : \mathbb{R}^n \to \mathbb{C}$ with derivatives of polynomial growth.

A special case is the following (see, e.g., [23, §5.5]).

Corollary 4.12 (Gaussian covariance identity)**.** *Let Y, Z be independent centered Gaussian vectors in $\mathbb{R}^n$ with covariance matrix Σ, and let*

$$Y'_t = t\,Y + \sqrt{1-t^2}\,Z$$

for $t \in [0,1]$. Then we have

$$\mathbf{E}[f(Y)g(Y)] - \mathbf{E}[f(Y)]\,\mathbf{E}[g(Y)] = \int_0^1 \mathbf{E}[\langle \nabla f(Y), \Sigma\,\nabla g(Y'_t)\rangle]\,dt$$

for any smooth $f, g : \mathbb{R}^n \to \mathbb{C}$ with derivatives of polynomial growth.

Proof. Let Y, Z, Z' be independent centered Gaussian vectors with covariance matrix Σ, and let $G = (Y, Y)$, $G' = (Z, Z')$, and $G_t = \sqrt{t}\,G + \sqrt{1-t}\,G'$. Then

$$\mathbf{E}[f(Y)g(Y)] - \mathbf{E}[f(Y)]\,\mathbf{E}[g(Y)] = \int_0^1 \frac{d}{dt}\mathbf{E}[H(G_t)]\,dt,$$

where $H(x, y) = f(x)g(y)$. The conclusion follows from Lemma 4.11 and the fact that $(\sqrt{t}\,Y + \sqrt{1-t}\,Z, \sqrt{t}\,Y + \sqrt{1-t}\,Z')$ is equidistributed with (Y, Y'_t). □

We finally recall the following [23, p. 41].

Lemma 4.13 (Gaussian concentration)**.** *Let Y be a standard Gaussian vector in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function. Then*

$$\mathbf{P}[f(Y) \geq \mathbf{E}f(Y) + t] \leq e^{-t^2/2L^2} \quad \textit{for all } t \geq 0.$$

It is instructive to spell out the application of Gaussian concentration to (2.1), which explains the significance of the parameter $\sigma_*(X)$.

Corollary 4.14. *Consider the model (2.1) with $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})$, and let $F : \mathrm{M}_d(\mathbb{C}) \to \mathbb{R}$ be L-Lipschitz with respect to the operator norm. Then*

$$\mathbf{P}[F(X) \geq \mathbf{E}F(X) + t] \leq e^{-t^2/2L^2\sigma_*(X)^2} \quad \textit{for all } t \geq 0.$$

If $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$, it suffices to assume F is L-Lipschitz on $\mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$.

Proof. We may write $F(X) = f(g_1, \ldots, g_n) := F(A_0 + \sum_i g_i A_i)$. Thus

$$\begin{aligned} |f(x) - f(y)| &\leq L\bigg\|\sum_i (x_i - y_i)A_i\bigg\| = L \sup_{\|v\|=\|w\|=1}\bigg|\sum_i (x_i - y_i)\langle v, A_i w\rangle\bigg| \\ &\leq L\sigma_*(X)\|x - y\| \end{aligned}$$

by Cauchy-Schwarz and the definition of $\sigma_*(X)$ (cf. Section 2.1). The conclusion follows by applying Lemma 4.13 to $f(g_1, \ldots, g_n)$. □

5. Spectral statistics

The next four sections are devoted to the proofs of the main results of this paper. In the present section, we begin by proving our bounds on the spectral statistics that were formulated in Section 2.2. These results illustrate the main proof technique of this paper in its simplest form. The support of the spectrum will be investigated in the next section using a more involved variant of the same method.

5.1. **The basic construction.** Throughout the proofs of our main results in Sections 2.1 and 2.2, we will fix $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ and let X and X_{free} be defined as in (2.1). (Where relevant, the extension to the non-self-adjoint case will be done at the end of the proof using Remark 2.6.)

Let $G_1^N, \ldots, G_n^N$ be independent standard Wigner matrices as in Definition 1.1, and let $D_1^N, \ldots, D_n^N$ be independent $N \times N$ diagonal matrices with i.i.d. standard Gaussians on the diagonal. We define for $q \in [0,1]$ the random matrix

$$X_q^N := A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes (\sqrt{q}\, D_i^N + \sqrt{1-q}\, G_i^N). \tag{5.1}$$

Note that $X_0^N = X^N$ as defined in (4.1). On the other hand, X_1^N is a block-diagonal matrix with i.i.d. copies of X on the diagonal. In particular, we have

$$\begin{aligned} \mathbf{E}[\mathrm{tr}\, h(X_1^N)] &= \mathbf{E}[\mathrm{tr}\, h(X)], \\ \mathbf{E}[\mathrm{tr}\, h(X_0^N)] &= \mathbf{E}[\mathrm{tr}\, h(X^N)] \end{aligned} \tag{5.2}$$

for any function $h : \mathbb{R} \to \mathbb{C}$. The basic idea behind our proofs is to interpolate between $\mathbf{E}[\mathrm{tr}\, h(X_1^N)]$ and $\mathbf{E}[\mathrm{tr}\, h(X_0^N)]$ using Lemma 4.11.

To simplify the expressions that will arise in the analysis, it will be convenient to define for $y = (y_{irs})_{1\le i\le n, 1\le s\le r\le N}$ the notation

$$X^N(y) := A_0 \otimes \mathbf{1} + \sum_{i=1}^n \sum_{1\le s\le r\le N} y_{irs}\, A_{irs}, \qquad A_{irs} := A_i \otimes E_{rs},$$

where E_{rs} are as defined in Section 3.1. Moreover, let Y, Z be centered Gaussian vectors all of whose entries $Y_{irs} = (D_i^N)_{rs}$ and $Z_{irs} = (G_i^N)_{rs}$ are independent with variances δ_{rs} and $\frac{1}{N}$, respectively. Then $X_q^N = X^N(\sqrt{q}\, Y + \sqrt{1-q}\, Z)$.

5.2. **Proof of Theorem 2.7.** In order to prove Theorem 2.7, we apply the above program to the moments. We begin with a simple computation.

Lemma 5.1. *For any $p \in \mathbb{N}$, we have*

$$\frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}] = p \sum_{k=0}^{2p-2} \sum_i \sum_{r\ge s} \left(\delta_{rs} - \frac{1}{N}\right) \mathbf{E}[\mathrm{tr}\, A_{irs}(X_q^N)^k A_{irs}(X_q^N)^{2p-2-k}].$$

Proof. Let $Y = (Y_{irs})_{i\in[n], r\ge s}$ and $Z = (Z_{irs})_{i\in[n], r\ge s}$ be the Gaussian vectors defined above. As both these vectors have independent entries, their covariance matrices Σ^Y and Σ^Z are diagonal with $\mathrm{Var}(Y_{irs}) = \delta_{rs}$ and $\mathrm{Var}(Z_{irs}) = \frac{1}{N}$. Applying Lemma 4.11 to the function $f(y) = \mathrm{tr}\, X^N(y)^{2p}$ therefore yields

$$\frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}] = \frac{1}{2} \sum_i \sum_{r\ge s} \left(\delta_{rs} - \frac{1}{N}\right) \mathbf{E}\left[\frac{\partial^2 f}{\partial y_{irs}^2}(\sqrt{q}\, Y + \sqrt{1-q}\, Z)\right].$$

The conclusion follows by a straightforward computation. ∎

As was explained in Section 1.4, we expect that the interpolation between X and X_{free} will be controlled only by the crossings in the moment formulae. This is however not immediately obvious from the expression in Lemma 5.1. To make this phenomenon visible, we need a simple lemma.

Lemma 5.2. $\mathbf{E}[h(X_q^N)] = \mathbf{E}[(\mathrm{id}\otimes\mathrm{tr})(h(X_q^N))]\otimes\mathbf{1}$ *for every* $h:\mathbb{R}\to\mathbb{C}$.

Proof. The distributions of D_i^N and G_i^N are invariant under conjugation by any signed permutation matrix. Therefore, if we let Π be an $N\times N$ signed permutation matrix chosen uniformly at random (independently of X_q^N), then

$$\mathbf{E}[h(X_q^N)] = \mathbf{E}[h((\mathbf{1}\otimes\Pi)^* X_q^N(\mathbf{1}\otimes\Pi))] = \mathbf{E}[(\mathbf{1}\otimes\Pi)^* h(X_q^N)(\mathbf{1}\otimes\Pi)].$$

It remains to note that $\mathbf{E}[(\mathbf{1}\otimes\Pi)^*M(\mathbf{1}\otimes\Pi)] = (\mathrm{id}\otimes\mathrm{tr})(M)\otimes\mathbf{1}$ for any matrix M (this is elementary when $M=A\otimes B$, and extends to general M by linearity). □

The key observation is now the following.

Corollary 5.3. *For any* $p\in\mathbb{N}$*, we have*

$$p\sum_{k=0}^{2p-2}\sum_i\sum_{r\ge s}\left(\delta_{rs}-\frac{1}{N}\right)\mathrm{tr}\,A_{irs}\mathbf{E}[(X_q^N)^k]A_{irs}\mathbf{E}[(X_q^N)^{2p-2-k}] = 0.$$

Proof. Note first that $E_{rs}^2 = E_{rr}^2 + E_{ss}^2$ for $r\ne s$. Thus

$$(A_i\otimes E_{rs})\mathbf{E}[(X_q^N)^k](A_i\otimes E_{rs}) = \\ (A_i\otimes E_{rr})\mathbf{E}[(X_q^N)^k](A_i\otimes E_{rr}) + (A_i\otimes E_{ss})\mathbf{E}[(X_q^N)^k](A_i\otimes E_{ss})$$

for $r\ne s$ by Lemma 5.2. Summing over $r>s$ yields

$$\begin{aligned}\frac{1}{N}\sum_{r>s}A_{irs}\mathbf{E}[(X_q^N)^k]A_{irs} &= \frac{1}{N}\sum_{r>s}(A_{irr}\mathbf{E}[(X_q^N)^k]A_{irr} + A_{iss}\mathbf{E}[(X_q^N)^k]A_{iss})\\ &= \left(1-\frac{1}{N}\right)\sum_r A_{irr}\mathbf{E}[(X_q^N)^k]A_{irr}.\end{aligned}$$

The conclusion follows readily. □

By combining Lemma 5.1 and Corollary 5.3, we can apply Corollary 4.12 to make crossings appear (the latter idea is already present in [37]). Recall that the parameters $w(X)$ and $w(X,X')$ were defined in Section 4.2.

Lemma 5.4. *For any* $p\in\mathbb{N}$*, we have*

$$\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]\right| \le \frac{4}{3}p^4\{qw(X_1^N)^4 + w(X_0^N,X_1^N)^4 + (1-q)w(X_0^N)^4\}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p-4}].$$

Proof. Recall that the random vectors Y,Z with $Y_{irs} = (D_i^N)_{rs}$ and $Z_{irs} = (G_i^N)_{rs}$ were defined in section 5.1. Let Y',Z' be independent copies of Y,Z, and define

$$X_{qt}^N = X^N\big(t\{\sqrt{q}\,Y + \sqrt{1-q}\,Z\} + \sqrt{1-t^2}\{\sqrt{q}\,Y' + \sqrt{1-q}\,Z'\}\big).$$

Note that the random vector $\sqrt{q}\,Y+\sqrt{1-q}\,Z$ has independent entries, so its covariance matrix Σ is diagonal with $\mathrm{Var}(\sqrt{q}\,Y_{irs}+\sqrt{1-q}\,Z_{irs}) = q\delta_{rs}+\frac{1-q}{N}$. We can therefore apply Corollary 4.12 to compute

$$\mathbf{E}[(X_q^N)^k_{ab}\,(X_q^N)^{2p-2-k}_{cd}] - \mathbf{E}[(X_q^N)^k_{ab}]\,\mathbf{E}[(X_q^N)^{2p-2-k}_{cd}] = \sum_{l=0}^{k-1}\sum_{m=0}^{2p-3-k}\sum_i\sum_{r\geq s}\left(q\delta_{rs}+\frac{1-q}{N}\right)\cdot \int_0^1 \mathbf{E}\big[\big((X_q^N)^l A_{irs}(X_q^N)^{k-1-l}\big)_{ab}\,\big((X_{qt}^N)^m A_{irs}(X_{qt}^N)^{2p-3-k-m}\big)_{cd}\big]\,dt.$$

Combining this identity with Lemma 5.1 and Corollary 5.3 yields

$$\begin{aligned}\frac{d}{dq}&\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]\\ &= p\sum_{k=0}^{2p-2}\sum_{i'}\sum_{r'\geq s'}\left(\delta_{r's'}-\frac{1}{N}\right)\mathbf{E}[\mathrm{tr}\,A_{i'r's'}(X_q^N)^k A_{i'r's'}(X_q^N)^{2p-2-k}]\\ &\quad - p\sum_{k=0}^{2p-2}\sum_{i'}\sum_{r'\geq s'}\left(\delta_{r's'}-\frac{1}{N}\right)\mathrm{tr}\,A_{i'r's'}\mathbf{E}[(X_q^N)^k]A_{i'r's'}\mathbf{E}[(X_q^N)^{2p-2-k}]\\ &= p\sum_{k=0}^{2p-2}\sum_{l=0}^{k-1}\sum_{m=0}^{2p-3-k}\int_0^1\sum_{i,i'}\sum_{r\geq s}\sum_{r'\geq s'}\left(q\delta_{rs}\delta_{r's'}+\frac{1-q}{N}\delta_{r's'}-\frac{q}{N}\delta_{rs}-\frac{1-q}{N^2}\right)\cdot\\ &\qquad \mathbf{E}[\mathrm{tr}\,A_{i'r's'}(X_q^N)^l A_{irs}(X_q^N)^{k-1-l}A_{i'r's'}(X_{qt}^N)^m A_{irs}(X_{qt}^N)^{2p-3-k-m}]\,dt.\end{aligned}$$

We can now apply Lemma 4.5 with

$$p_1=\frac{2p-4}{l},\qquad p_2=\frac{2p-4}{k-1-l},\qquad p_3=\frac{2p-4}{m},\qquad p_4=\frac{2p-4}{2p-3-k-m},$$

to bound, for example,

$$\left|\sum_{i,i'}\sum_{r\geq s}\sum_{r'\geq s'}\frac{q}{N}\delta_{rs}\,\mathbf{E}[\mathrm{tr}\,A_{i'r's'}(X_q^N)^l A_{irs}(X_q^N)^{k-1-l}A_{i'r's'}(X_{qt}^N)^m\cdot A_{irs}(X_{qt}^N)^{2p-3-k-m}]\right| \leq q w(X_0^N,X_1^N)^4\mathbf{E}[\mathrm{tr}(X_q^N)^{2p-4}],$$

where we used that $\mathrm{Var}(Y_{irs})=\delta_{rs}$, $\mathrm{Var}(Z_{irs})=\frac{1}{N}$, and that X_q^N and X_{qt}^N are equidistributed. The remaining three terms in the integral can be bounded analogously. To conclude, it remains to note that $\sum_{k=0}^{2p-2}k(2p-2-k)=\binom{2p-1}{3}\leq\frac{4}{3}p^3$. □

Before we can complete the proof of Theorem 2.7, we must compute the matrix parameters associated to X_q^N.

Lemma 5.5. *For every q,N, we have*

$$\sigma(X_q^N)=\sigma(X),\qquad v(X_1^N)=v(X),\qquad v(X_0^N)=\frac{v(X)}{\sqrt{N}}.$$

Proof. As $\mathbf{E}[(D_i^N)^2]=\mathbf{E}[(G_i^N)^2]=\mathbf{1}$, we have $\mathbf{E}[(X_q^N-\mathbf{E}X_q^N)^2]=\sum_i A_i^2\otimes\mathbf{1}$ and thus $\sigma(X_q^N)^2=\|\mathbf{E}[(X_q^N-\mathbf{E}X_q^N)^2]\|=\sigma(X)^2$.

Next, note that X_1^N is a block-diagonal matrix with i.i.d. copies of X on the diagonal, so $v(X_1^N)^2=\|\mathrm{Cov}(X_1^N)\|=\|\mathrm{Cov}(X)\|=v(X)^2$. On the other hand, X_0^N is a symmetric block matrix whose blocks on and above the diagonal are i.i.d. copies of $N^{-\frac{1}{2}}X$, so $v(X_0^N)^2=\|\mathrm{Cov}(X_0^N)\|=N^{-1}\|\mathrm{Cov}(X)\|=N^{-1}v(X)^2$. □

We can now conclude the proof.

Proof of Theorem 2.7. Assume first that $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ are self-adjoint. Applying Lemma 5.4, the chain rule, and Proposition 4.6 yields

$$\begin{aligned}\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]^{\frac{2}{p}}\right| &= \frac{2}{p}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]^{\frac{2}{p}-1}\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]\right| \\ &\le \frac{8}{3}p^3\{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\} \\ &\le \frac{8}{3}p^3\{q\tilde{v}(X_1^N)^4 + \tilde{v}(X_0^N)^2\tilde{v}(X_1^N)^2 + (1-q)\tilde{v}(X_0^N)^4\},\end{aligned}$$

where we used that $\mathbf{E}[\mathrm{tr}(X_q^N)^{2p-4}] \le \mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]^{1-\frac{2}{p}}$ by Hölder's inequality. Thus

$$\begin{aligned}|\mathbf{E}[\mathrm{tr}\,X^{2p}]^{\frac{1}{2p}} - \mathbf{E}[\mathrm{tr}(X^N)^{2p}]^{\frac{1}{2p}}| &\le |\mathbf{E}[\mathrm{tr}\,X^{2p}]^{\frac{2}{p}} - \mathbf{E}[\mathrm{tr}(X^N)^{2p}]^{\frac{2}{p}}|^{\frac{1}{4}} \\ &= \left|\int_0^1 \frac{d}{dq}\mathbf{E}[\mathrm{tr}(X_q^N)^{2p}]^{\frac{2}{p}}\,dq\right|^{\frac{1}{4}} \le \left(\frac{4}{3}\right)^{\frac{1}{4}} p^{\frac{3}{4}}\{\tilde{v}(X_1^N)^2 + \tilde{v}(X_0^N)^2\}^{\frac{1}{2}},\end{aligned}$$

where we used $x - y = (x^4 - y^4 + y^4)^{\frac{1}{4}} - y \le (x^4 - y^4)^{\frac{1}{4}}$ for $x \ge y \ge 0$ and (5.2). But note that Lemma 5.5 implies $\tilde{v}(X_1^N) = \tilde{v}(X)$ and $\tilde{v}(X_0^N) = N^{-\frac{1}{4}}\tilde{v}(X)$. We may therefore let $N \to \infty$ in the above inequality and use Corollary 4.4 to obtain

$$|\mathbf{E}[\mathrm{tr}\,X^{2p}]^{\frac{1}{2p}} - (\mathrm{tr}\otimes\tau)(X_{\mathrm{free}}^{2p})^{\frac{1}{2p}}| \le \left(\frac{4}{3}\right)^{\frac{1}{4}} p^{\frac{3}{4}}\tilde{v}(X).$$

Finally, we extend the conclusion to non-self-adjoint $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})$ by applying the above inequality to the self-adjoint model $\breve{X}$ defined in Remark 2.6. As $\mathbf{E}[\mathrm{tr}\,\breve{X}^{2p}] = \mathbf{E}[\mathrm{tr}\,|X|^{2p}]$ and $(\mathrm{tr}\otimes\tau)(\breve{X}_{\mathrm{free}}^{2p}) = (\mathrm{tr}\otimes\tau)(|X_{\mathrm{free}}|^{2p})$ by (4.3), and as $\tilde{v}(\breve{X}) \le 2^{\frac{1}{4}}\tilde{v}(X)$, the conclusion follows readily (using $(\frac{8}{3})^{\frac{1}{4}} \le 2$). □

Remark 5.6. When $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ are self-adjoint, we may obtain a slightly better bound in the proof of Theorem 2.7 by neglecting to apply Proposition 4.6 to $w(X_1^N)$. In this case, the parameter $\tilde{v}(X)$ in the final bound is replaced by $\sup_N w(X_1^N)$. The analogous improvement is possible for most results of this paper. However, as $\sup_N w(X_1^N)$ is very difficult to compute in any concrete situation, we have formulated our main results in terms of the computable quantity $\tilde{v}(X)$.

5.3. **Proof of Theorem 2.8.** Once the basic method of proof has been understood, it may be readily adapted to control spectral statistics other than the moments. We presently adapt the method of the previous section to the matrix-valued Stieltjes transform. Note that Theorem 2.8 assumes $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$.

Lemma 5.7. *For any $Z \in \mathrm{M}_d(\mathbb{C})$, $\mathrm{Im}\,Z > 0$ and $M \in \mathrm{M}_d(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$, we have*

$$\begin{aligned}&\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,M(\tilde{Z} - X_q^N)^{-1}] = \\ &\qquad \sum_i \sum_{r\ge s}\left(\delta_{rs} - \frac{1}{N}\right)\mathbf{E}[\mathrm{tr}\,A_{irs}(\tilde{Z} - X_q^N)^{-1}A_{irs}(\tilde{Z} - X_q^N)^{-1}M(\tilde{Z} - X_q^N)^{-1}]\end{aligned}$$

and

$$\sum_i \sum_{r\ge s}\left(\delta_{rs} - \frac{1}{N}\right)\mathrm{tr}\,A_{irs}\mathbf{E}[(\tilde{Z} - X_q^N)^{-1}]A_{irs}\mathbf{E}[(\tilde{Z} - X_q^N)^{-1}M(\tilde{Z} - X_q^N)^{-1}] = 0,$$

where we defined $\tilde{Z} = Z \otimes \mathbf{1} \in \mathrm{M}_d(\mathbb{C}) \otimes \mathrm{M}_N(\mathbb{C})$.

Proof. The first identity follows from Lemma 4.11 with $f(y) = \mathrm{tr}\, M(\tilde{Z} - X^N(y))^{-1}$. The second identity follows as $\mathbf{E}[(\tilde{Z} - X_q^N)^{-1}] = \mathbf{E}[(\mathrm{id} \otimes \mathrm{tr})(\tilde{Z} - X_q^N)^{-1}] \otimes \mathbf{1}$ holds by precisely the same proof as that of Lemma 5.2. □

We can now proceed as in Lemma 5.4.

Lemma 5.8. *For any* $Z \in \mathrm{M}_d(\mathbb{C})$, $\mathrm{Im}\, Z > 0$ *we have*

$$\left\| \frac{d}{dq} \mathbf{E}[(Z \otimes \mathbf{1} - X_q^N)^{-1}] \right\| \le 2\|(\mathrm{Im}\, Z)^{-5}\| \{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\}.$$

Proof. Define X_{qt}^N as in the proof of Lemma 5.4, and denote $R := (Z \otimes \mathbf{1} - X_q^N)^{-1}$ and $R_t := (Z \otimes \mathbf{1} - X_{qt}^N)^{-1}$ for simplicity. Corollary 4.12 and Lemma 5.7 yield

$$\begin{aligned}
&\frac{d}{dq} \mathbf{E}[\mathrm{tr}\, M(Z \otimes \mathbf{1} - X_q^N)^{-1}] = \\
&\int_0^1 \sum_{i,i'} \sum_{r \ge s} \sum_{r' \ge s'} \left(q\delta_{rs}\delta_{r's'} + \frac{1-q}{N}\delta_{r's'} - \frac{q}{N}\delta_{rs} - \frac{1-q}{N^2} \right) \cdot \\
&\qquad \big\{ \mathbf{E}[\mathrm{tr}\, A_{i'r's'} R A_{irs} R A_{i'r's'} R_t A_{irs} R_t M R_t] \\
&\qquad\quad + \mathbf{E}[\mathrm{tr}\, A_{i'r's'} R A_{irs} R A_{i'r's'} R_t M R_t A_{irs} R_t] \big\}\, dt.
\end{aligned}$$

Now apply Lemma 4.5 with $p_1 = p_2 = p_3 = \infty$ and $p_4 = 1$ to the first expectation in the integral, and with $p_1 = p_2 = p_4 = \infty$ and $p_3 = 1$ to the second expectation. This yields, in the same manner as in the proof of Lemma 5.4, that

$$\begin{aligned}
&\left| \frac{d}{dq} \mathbf{E}[\mathrm{tr}\, M(Z \otimes \mathbf{1} - X_q^N)^{-1}] \right| \\
&\qquad \le 2\{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\} \| \|R\| \|_\infty^3 \mathbf{E}[\mathrm{tr}\, |RMR|].
\end{aligned}$$

But as $\|R\| \le \|(\mathrm{Im}\, Z)^{-1}\|$ (see, e.g., [19, Lemma 3.1]), we obtain

$$\begin{aligned}
&\left| \mathrm{tr}\, M \frac{d}{dq} \mathbf{E}[(Z \otimes \mathbf{1} - X_q^N)^{-1}] \right| \\
&\qquad \le 2\|(\mathrm{Im}\, Z)^{-5}\| \{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\}\, \mathrm{tr}\, |M|.
\end{aligned}$$

The conclusion follows by taking the supremum over all M with $\mathrm{tr}\, |M| \le 1$. □

Integrating the above differential inequality yields the following.

Lemma 5.9. *For any* $Z \in \mathrm{M}_d(\mathbb{C})$, $\mathrm{Im}\, Z > 0$ *we have*

$$\|\mathbf{E}[(Z - X)^{-1}] - \mathbf{E}[(\mathrm{id} \otimes \mathrm{tr})(Z \otimes \mathbf{1} - X^N)^{-1}]\| \le (1 + N^{-\frac{1}{2}})^2 \tilde{v}(X)^4 \|(\mathrm{Im}\, Z)^{-5}\|.$$

Proof. Integrating Lemma 5.8 and using Proposition 4.6 yields

$$\|\mathbf{E}[(Z \otimes \mathbf{1} - X_1^N)^{-1}] - \mathbf{E}[(Z \otimes \mathbf{1} - X_0^N)^{-1}]\| \le \{\tilde{v}(X_1^N)^2 + \tilde{v}(X_0^N)^2\}^2 \|(\mathrm{Im}\, Z)^{-5}\|.$$

As $X_0^N = X^N$, we have $\mathbf{E}[(Z \otimes \mathbf{1} - X_0^N)^{-1}] = \mathbf{E}[(\mathrm{id} \otimes \mathrm{tr})(Z \otimes \mathbf{1} - X^N)^{-1}] \otimes \mathbf{1}$ as in Lemma 5.2. Similarly, as X_1^N is block-diagonal with i.i.d. copies of X on the diagonal, we have $\mathbf{E}[(Z \otimes \mathbf{1} - X_1^N)^{-1}] = \mathbf{E}[(Z - X)^{-1}] \otimes \mathbf{1}$ as in Lemma 5.2. The conclusion follows readily from these observations and Lemma 5.5. □

It remains to take the limit $N \to \infty$ in Lemma 5.9. While Corollary 4.4 does not apply directly here, its proof may be readily extended to the present setting.

Lemma 5.10. *For any $Z \in \mathrm{M}_d(\mathbb{C})$, $\mathrm{Im}\, Z > 0$ we have*

$$\lim_{N\to\infty} \|\mathbf{E}[(\mathrm{id}\otimes\mathrm{tr})(Z\otimes\mathbf{1} - X^N)^{-1}] - (\mathrm{id}\otimes\tau)(Z\otimes\mathbf{1} - X_{\mathrm{free}})^{-1}\| = 0.$$

Proof. As we aim to establish convergence as $N \to \infty$ in $\mathrm{M}_d(\mathbb{C})$ with a fixed finite dimension d, it suffices to show that

$$\lim_{N\to\infty} \langle v, \{\mathbf{E}[(\mathrm{id}\otimes\mathrm{tr})(Z\otimes\mathbf{1} - X^N)^{-1}] - (\mathrm{id}\otimes\tau)(Z\otimes\mathbf{1} - X_{\mathrm{free}})^{-1}\}\, v\rangle = 0$$

for all $v \in \mathbb{C}^d$ with $\|v\| = 1$. Moreover, if we define

$$\tilde X^N := (\mathrm{Im}\, Z\otimes\mathbf{1})^{-1/2}\{X^N - \mathrm{Re}\, Z\otimes\mathbf{1}\}(\mathrm{Im}\, Z\otimes\mathbf{1})^{-1/2},$$
$$\tilde X_{\mathrm{free}} := (\mathrm{Im}\, Z\otimes\mathbf{1})^{-1/2}\{X_{\mathrm{free}} - \mathrm{Re}\, Z\otimes\mathbf{1}\}(\mathrm{Im}\, Z\otimes\mathbf{1})^{-1/2}$$

where $\mathrm{Re}\, Z := \frac{1}{2}(Z + Z^*)$, it clearly suffices to show that

$$\lim_{N\to\infty} \langle v, \{\mathbf{E}[(\mathrm{id}\otimes\mathrm{tr})(i\mathbf{1} - \tilde X^N)^{-1}] - (\mathrm{id}\otimes\tau)(i\mathbf{1} - \tilde X_{\mathrm{free}})^{-1}\}\, v\rangle = 0$$

for all $v \in \mathbb{C}^d$ with $\|v\| = 1$. By the spectral theorem, there are probability measures μ_N, μ (which depend on the choice of v) so that

$$\int h\, d\mu_N = \langle v, \mathbf{E}[(\mathrm{id}\otimes\mathrm{tr})(h(\tilde X^N))]\, v\rangle, \qquad \int h\, d\mu = \langle v, (\mathrm{id}\otimes\tau)(h(\tilde X_{\mathrm{free}}))\, v\rangle$$

for $h : \mathbb{R} \to \mathbb{C}$. Theorem 4.3 yields $\int x^p\, d\mu_N \to \int x^p\, d\mu$ for $p \in \mathbb{N}$ as in the proof of Corollary 4.4. As $\|\tilde X_{\mathrm{free}}\| < \infty$, the measure μ has bounded support. Thus moment convergence implies weak convergence [27, p. 116], concluding the proof. □

Proof of Theorem 2.8. The conclusion follows immediately by taking $N \to \infty$ in Lemma 5.9 and using Lemma 5.10. □

5.4. **Proof of Corollary 2.9.** The deduction of Corollary 2.9 from Theorem 2.8 follows by applying general facts about Stieltjes transforms that may be found in [19, §6]. For convenience, we formulate a general statement.

Lemma 5.11. *Let μ, ν be probability measures on $\mathbb{R}$ with Stieltjes transforms*

$$s_\mu(z) := \int \frac{1}{z-x}\,\mu(dx), \qquad s_\nu(z) := \int \frac{1}{z-x}\,\nu(dx).$$

Suppose that

$$|s_\mu(z) - s_\nu(z)| \le \frac{K}{(\mathrm{Im}\, z)^p}$$

for some $K \ge 0$, $p \in \mathbb{N}$, and all $z \in \mathbb{C}$ with $\mathrm{Im}\, z > 0$. Then

$$\left|\int h\, d\mu - \int h\, d\nu\right| \le \frac{(\sqrt{2})^{p+1}K}{p!\pi}\int_\infty^\infty \left|\left(1 + \frac{d}{dx}\right)^{p+1} h(x)\right| dx \lesssim K\|h\|_{W^{p+1,1}(\mathbb{R})}$$

for every $h \in W^{p+1,1}(\mathbb{R})$.

Proof. Let $h \in C_c^\infty(\mathbb{R})$. Following *verbatim* the proof of [19, Theorem 6.2] yields

$$\left|\int h\, d\mu - \int h\, d\nu\right| \le \frac{1}{\pi}\limsup_{y\downarrow 0}\int_\infty^\infty \left|\left(1 + \frac{d}{dx}\right)^{p+1} h(x)\right| |I_{p+1}(x+iy)|\, dx$$

with

$$|I_{p+1}(z)| \le \frac{1}{p!}\int_0^\infty \frac{K}{(\mathrm{Im}\, z + t)^p}(\sqrt{2}t)^p e^{-t}\sqrt{2}\, dt \le \frac{(\sqrt{2})^{p+1}K}{p!}.$$

That the integral may be bounded up to a universal constant by the Sobolev norm $\|h\|_{W^{p+1,1}(\mathbb{R})}$ follows as $\binom{p+1}{k}\frac{(\sqrt{2})^{p+1}}{p!} \lesssim 1$ for all $0 \le k \le p+1$. The conclusion finally extends to general $h \in W^{p+1,1}(\mathbb{R})$ by routine approximation arguments. □

We can now conclude the proof.

Proof of Corollary 2.9. Theorem 2.8 implies

$$|\mathbf{E}[\mathrm{tr}(z\mathbf{1}-X)^{-1}] - (\mathrm{tr}\otimes\tau)(z\mathbf{1}-X_{\mathrm{free}})^{-1}| \le \frac{\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5}$$

for all $z \in \mathbb{C}$ with $\mathrm{Im}\, z > 0$. Applying Lemma 5.11 with $p = 5$ to the spectral distributions of X and X_{free} immediately yields the conclusion. □

6. Concentration of the spectrum

The aim of this section is to prove our main results on the support of the spectrum that were formulated in Section 2.1. The general scheme of proof is the same as in the previous section, but some new ingredients are needed here.

6.1. **Moments of the resolvent.** The proof of Theorem 2.1 is based on an analysis of large moments of the resolvent $\mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X|^{-2p}]$. In the present section, we will prove an analogue of Theorem 2.8 for these higher moments.

Theorem 6.1. *Let $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$. Then we have*

$$|\mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X|^{-2p}]^{\frac{1}{2p}} - (\mathrm{tr}\otimes\tau)(|z\mathbf{1}-X_{\mathrm{free}}|^{-2p})^{\frac{1}{2p}}| \le \frac{(p+2)^3}{3}\frac{\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5}$$

for every $p \in \mathbb{N}$ and $z \in \mathbb{C}$, $\mathrm{Im}\, z > 0$.

The proof of Theorem 6.1 is similar to that of Theorems 2.7 and 2.8. Throughout this section, we adopt without further comment the constructions and notation of Section 5.1. In particular, X_q^N is defined as in (5.1).

Lemma 6.2. *For any $p \in \mathbb{N}$ and $z \in \mathbb{C}$, $\mathrm{Im}\, z > 0$, we have*

$$\begin{aligned}\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X_q^N|^{-2p}] = p\sum_i\sum_{r\ge s}\left(\delta_{rs}-\frac{1}{N}\right)\cdot \\ \Bigg\{\sum_{k=0}^{p}\mathrm{Re}\,\mathbf{E}[\mathrm{tr}\,A_{irs}(z\mathbf{1}-X_q^N)^{-k-1}A_{irs}(z\mathbf{1}-X_q^N)^{-p-1+k}(\bar{z}\mathbf{1}-X_q^N)^{-p}] \\ +\sum_{k=0}^{p-1}\mathrm{Re}\,\mathbf{E}[\mathrm{tr}\,A_{irs}(z\mathbf{1}-X_q^N)^{-p-1}(\bar{z}\mathbf{1}-X_q^N)^{-k-1}A_{irs}(\bar{z}\mathbf{1}-X_q^N)^{-p+k}]\Bigg\}\end{aligned}$$

and

$$\begin{aligned}0 = p\sum_i\sum_{r\ge s}\left(\delta_{rs}-\frac{1}{N}\right)\cdot \\ \Bigg\{\sum_{k=0}^{p}\mathrm{Re}\ \mathrm{tr}\,A_{irs}\mathbf{E}[(z\mathbf{1}-X_q^N)^{-k-1}]A_{irs}\mathbf{E}[(z\mathbf{1}-X_q^N)^{-p-1+k}(\bar{z}\mathbf{1}-X_q^N)^{-p}] \\ +\sum_{k=0}^{p-1}\mathrm{Re}\ \mathrm{tr}\,A_{irs}\mathbf{E}[(z\mathbf{1}-X_q^N)^{-p-1}(\bar{z}\mathbf{1}-X_q^N)^{-k-1}]A_{irs}\mathbf{E}[(\bar{z}\mathbf{1}-X_q^N)^{-p+k}]\Bigg\}.\end{aligned}$$

Proof. The first identity follows by applying Lemma 4.11 to the function

$$f(y) = \mathrm{tr}\,|z\mathbf{1} - X^N(y)|^{-2p} = \mathrm{tr}[(z\mathbf{1} - X^N(y))^{-p}(\overline{z}\mathbf{1} - X^N(y))^{-p}].$$

The second identity follows by applying Lemma 5.2. □

We can now proceed as in Lemma 5.4.

Lemma 6.3. *For any $p \in \mathbb{N}$ and $z \in \mathbb{C}$, $\mathrm{Im}\, z > 0$, we have*

$$\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p}]\right|$$
$$\leq \frac{4}{3}p(p+2)^3\{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p-4}].$$

Proof. Define X_{qt}^N as in the proof of Lemma 5.4, and denote $R := (z\mathbf{1} - X_q^N)^{-1}$ and $R_t := (z\mathbf{1} - X_{qt}^N)^{-1}$. Applying Corollary 4.12 and Lemma 6.2 yields

$$\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p}] =$$
$$p\,\mathrm{Re}\int_0^1 \sum_{i,i'}\sum_{r\geq s}\sum_{r'\geq s'}\left(q\delta_{rs}\delta_{r's'} + \frac{1-q}{N}\delta_{rs} - \frac{q}{N}\delta_{r's'} - \frac{1-q}{N^2}\right)\cdot$$
$$\left\{\sum_{k=0}^{p-1}\sum_{l=0}^{p}\sum_{m=0}^{p-k-1}\mathbf{E}[\mathrm{tr}\,A_{irs}R^{l+1}A_{i'r's'}R^{p-l+1}R^{*(k+1)}A_{irs}R_t^{*(m+1)}A_{i'r's'}R_t^{*(p-k-m)}] +\right.$$
$$\sum_{k=0}^{p-1}\sum_{l=0}^{k}\sum_{m=0}^{p-k-1}\mathbf{E}[\mathrm{tr}\,A_{irs}R^{p+1}R^{*(l+1)}A_{i'r's'}R^{*(k-l+1)}A_{irs}R_t^{*(m+1)}A_{i'r's'}R_t^{*(p-k-m)}] +$$
$$\sum_{k=0}^{p}\sum_{l=0}^{k}\sum_{m=0}^{p-k}\mathbf{E}[\mathrm{tr}\,A_{irs}R^{l+1}A_{i'r's'}R^{k-l+1}A_{irs}R_t^{m+1}A_{i'r's'}R_t^{p-k-m+1}R_t^{*p}] +$$
$$\left.\sum_{k=0}^{p}\sum_{l=0}^{k}\sum_{m=0}^{p-1}\mathbf{E}[\mathrm{tr}\,A_{irs}R^{l+1}A_{i'r's'}R^{k-l+1}A_{irs}R_t^{p+1-k}R_t^{*(m+1)}A_{i'r's'}R_t^{*(p-m)}]\right\}dt.$$

We can now apply Lemma 4.5 as in the proof of Lemma 5.4 to bound

$$\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p}]\right|$$
$$\leq p\binom{2p+3}{3}\{qw(X_1^N)^4 + w(X_0^N, X_1^N)^4 + (1-q)w(X_0^N)^4\}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p-4}].$$

The conclusion follows using $\binom{2p+3}{3} \leq \frac{4}{3}(p+2)^3$. □

We can now complete the proof.

Proof of Theorem 6.1. Lemma 6.3, the chain rule, and Proposition 4.6 yield

$$\left|\frac{d}{dq}\mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X_q^N|^{-2p}]^{\frac{1}{2p}}\right| \leq \frac{2}{3}\frac{(p+2)^3}{(\mathrm{Im}\,z)^5}\{q\tilde{v}(X_1^N)^4+\tilde{v}(X_0^N)^2\tilde{v}(X_1^N)^2+(1-q)\tilde{v}(X_0^N)^4\},$$

where we used that

$$\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p-4}] \leq \frac{\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p+1}]}{(\mathrm{Im}\,z)^5} \leq \frac{\mathbf{E}[\mathrm{tr}\,|z\mathbf{1} - X_q^N|^{-2p}]^{1-\frac{1}{2p}}}{(\mathrm{Im}\,z)^5}$$

using $\||z\mathbf{1}-X_q^N|^{-1}\| \le (\mathrm{Im}\, z)^{-1}$ and Hölder's inequality. Integrating yields

$$|\mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X|^{-2p}]^{\frac{1}{2p}} - \mathbf{E}[\mathrm{tr}\,|z\mathbf{1}-X^N|^{-2p}]^{\frac{1}{2p}}| \le \frac{(1+N^{-1})^2}{3}\frac{(p+2)^3\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5}$$

using (5.2) and Lemma 5.5. It remains to let $N\to\infty$ using Corollary 4.4. □

6.2. **Proof of Theorem 2.1.** The basic observation behind the proof is the following. For any $D\subseteq\mathbb{C}$ and $z\in\mathbb{C}$, denote $d(z,D) := \inf_{z'\in D}|z-z'|$. Then

$$\|(z\mathbf{1}-X)^{-1}\| = \frac{1}{d(z,\mathrm{sp}(X))}, \tag{6.1}$$

and analogously for X_{free}. The following device will enable us to deduce concentration of the spectrum from resolvent inequalities.

Lemma 6.4. *Let $K,L\ge 0$, and let A,B be self-adjoint operators such that*

$$\|(z\mathbf{1}-A)^{-1}\| \le C\|(z\mathbf{1}-B)^{-1}\| + \frac{K}{(\mathrm{Im}\, z)^5} + \frac{L}{(\mathrm{Im}\, z)^2}$$

for all $z=\lambda+i\varepsilon$ with $\lambda\in\mathrm{sp}(A)$ and $\varepsilon=(4K)^{\frac14}\vee 4L$. Then

$$\mathrm{sp}(A)\subseteq\mathrm{sp}(B)+2C\varepsilon[-1,1].$$

Proof. By (6.1), the assumption states that

$$\frac{1}{\varepsilon} \le \frac{C}{\sqrt{\varepsilon^2+d(\lambda,\mathrm{sp}(B))^2}} + \frac{K}{\varepsilon^5} + \frac{L}{\varepsilon^2} \qquad \text{for all } \lambda\in\mathrm{sp}(A).$$

If $d(\lambda,\mathrm{sp}(B)) > 2C\varepsilon$, we would have $\frac12 < \frac{K}{\varepsilon^4}+\frac{L}{\varepsilon}\le\frac12$, which entails a contradiction. Thus we have shown that $d(\lambda,\mathrm{sp}(B))\le 2C\varepsilon$ for all $\lambda\in\mathrm{sp}(A)$. □

Our aim is now to show that the condition of Lemma 6.4 holds with high probability for $A=X$ and $B=X_{\mathrm{free}}$. To this end, we begin by showing that the relevant condition holds with high probability for a given $z\in\mathbb{C}$.

Lemma 6.5. *Fix $z\in\mathbb{C}$ with $\mathrm{Im}\, z>0$. Then*

$$\mathbf{P}\left[\|(z\mathbf{1}-X)^{-1}\| \ge \sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5} + \frac{\sigma_*(X)}{(\mathrm{Im}\, z)^2}t\right] \le e^{-\frac{t^2}{2}}$$

for all $t\ge 0$.

Proof. Using that $\mathrm{tr}\,|M|\ge\frac1d\|M\|$ for every $M\in\mathrm{M}_d(\mathbb{C})$, Theorem 6.1 yields

$$d^{-\frac{1}{2p}}\mathbf{E}\|(z\mathbf{1}-X)^{-1}\| \le \|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \frac{(p+2)^3}{3}\frac{\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5}$$

for every $p\in\mathbb{N}$. Choosing $p=\lceil\log d\rceil$ yields

$$\mathbf{E}\|(z\mathbf{1}-X)^{-1}\| \le \sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\mathrm{Im}\, z)^5}.$$

It remains to note that $F(X)=\|(z\mathbf{1}-X)^{-1}\|$ satisfies

$$|F(X)-F(Y)| \le \|(z\mathbf{1}-X)^{-1}(X-Y)(z\mathbf{1}-Y)^{-1}\| \le \frac{\|X-Y\|}{(\mathrm{Im}\, z)^2} \tag{6.2}$$

for $X,Y\in\mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$, so that the conclusion follows from Corollary 4.14. □

We must now show that $\|(z\mathbf{1}-X)^{-1}\|$ is small with high probability simultaneously for all $z=\lambda+i\varepsilon$ with $\lambda\in\mathrm{sp}(X)$. To create the requisite uniformity in z, we first need a crude *a priori* bound on the spectrum of X.

Lemma 6.6. *For any $t\ge 0$, we have*

$$\mathbf{P}[\mathrm{sp}(X)\subseteq\mathrm{sp}(A_0)+\sigma_*(X)\{d+t\}[-1,1]]\ge 1-e^{-\frac{t^2}{2}}.$$

Proof. By Weyl's inequality, we have $|\lambda_i(X)-\lambda_i(A_0)|\le\|X-A_0\|$ for every i, where $\lambda_i(X)$ denotes the ith largest eigenvalue of X. Thus

$$\mathrm{sp}(X)\subseteq\mathrm{sp}(A_0)+\|X-A_0\|[-1,1].$$

By Cauchy-Schwarz, we can crudely bound

$$\|X-A_0\|=\sup_{\|v\|=\|w\|=1}\left|\sum_{i=1}^n g_i\langle v,A_iw\rangle\right|\le\sigma_*(X)\|g\|.$$

Thus we have shown

$$\mathbf{P}[\mathrm{sp}(X)\subseteq\mathrm{sp}(A_0)+\sigma_*(X)\{d+t\}[-1,1]]\ge\mathbf{P}[\|g\|\le d+t].$$

But note that the argument in the proof of Lemma 4.7 shows that we may assume $n\le d^2$ without loss of generality. Thus $\mathbf{E}\|g\|\le\sqrt{n}\le d$. It remains to note that

$$\mathbf{P}[\|g\|\ge d+t]\le\mathbf{P}[\|g\|\ge\mathbf{E}\|g\|+t]\le e^{-\frac{t^2}{2}}$$

by Lemma 4.13. □

We are now ready to prove a uniform analogue of Lemma 6.5.

Lemma 6.7. *Fix $\varepsilon>0$. Then*

$$\mathbf{P}\Bigg[\|(z\mathbf{1}-X)^{-1}\|\le\sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\|+\sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\mathrm{Im}\,z)^5} \\ +(\sqrt{e}+2)\frac{\sigma_*(X)}{(\mathrm{Im}\,z)^2}(4\sqrt{\log d}+t)\quad\text{for all } z\in\mathrm{sp}(X)+i\varepsilon\Bigg]\ge 1-e^{-\frac{t^2}{2}}$$

for all $t\ge 0$.

Proof. Define the (nonrandom) set

$$\Omega_t:=\mathrm{sp}(A_0)+\sigma_*(X)\{d+t\}[-1,1]\subset\mathbb{R}.$$

As A_0 has at most d distinct eigenvalues, Ω_t is the union of at most d intervals of length $2\sigma_*(X)\{d+t\}$. We can therefore find $\mathcal{N}_t\subset\Omega_t$ of cardinality $|\mathcal{N}_t|\le\frac{2d(d+t)}{t}$ such that each $\lambda\in\Omega_t$ satisfies $d(\lambda,\mathcal{N}_t)\le\sigma_*(X)t$.

Now note that we can estimate as in (6.2)

$$\big|\|(z\mathbf{1}-X)^{-1}\|-\|(z'\mathbf{1}-X)^{-1}\|\big|\le\frac{|z-z'|}{\mathrm{Im}\,z\cdot\mathrm{Im}\,z'},$$

and similarly for X_{free}. We therefore obtain

$$\mathbf{P}\bigg[\|(z\mathbf{1}-X)^{-1}\| \le \sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\operatorname{Im} z)^5} + (\sqrt{e}+2)\frac{\sigma_*(X)}{(\operatorname{Im} z)^2}t \quad \text{for all } z \in \Omega_t + i\varepsilon\bigg] \ge$$
$$\mathbf{P}\bigg[\|(z\mathbf{1}-X)^{-1}\| \le \sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\operatorname{Im} z)^5} + \frac{\sigma_*(X)}{(\operatorname{Im} z)^2}t \quad \text{for all } z \in \mathcal{N}_t + i\varepsilon\bigg] \ge 1 - |\mathcal{N}_t|e^{-\frac{t^2}{2}},$$

where we used that $\operatorname{Im} z = \operatorname{Im} z' = \varepsilon$ for $z, z' \in \Omega_t + i\varepsilon$ in the first inequality, and we used the union bound and Lemma 6.5 in the second inequality. In particular,

$$\mathbf{P}\bigg[\|(z\mathbf{1}-X)^{-1}\| \le \sqrt{e}\|(z\mathbf{1}-X_{\mathrm{free}})^{-1}\| + \sqrt{e}\frac{(\log d+3)^3}{3}\frac{\tilde{v}(X)^4}{(\operatorname{Im} z)^5} + (\sqrt{e}+2)\frac{\sigma_*(X)}{(\operatorname{Im} z)^2}t \quad \text{for all } z \in \operatorname{sp}(X) + i\varepsilon\bigg] \ge 1 - (|\mathcal{N}_t|+1)e^{-\frac{t^2}{2}}$$

by Lemma 6.6. It remains to note that $(|\mathcal{N}_{t+a}|+1)e^{-\frac{(t+a)^2}{2}} \le e^{-\frac{t^2}{2}}$ if we choose $a = 4\sqrt{\log d}$ (recalling the standing assumption $d \ge 2$). ☐

The proof of Theorem 2.1 now follows readily.

Proof of Theorem 2.1. Combining Lemmas 6.4 and 6.7 yields

$$\mathbf{P}\big[\operatorname{sp}(X) \subseteq \operatorname{sp}(X_{\mathrm{free}}) + C\{\tilde{v}(X)(\log d)^{\frac{3}{4}} + \sigma_*(X)(\sqrt{\log d}+t)\}[-1,1]\big] \ge 1 - e^{-t^2}$$

for all $t \ge 0$, where C is a universal constant. It remains to note that we can estimate $\sigma_*(X)\sqrt{\log d} \lesssim \tilde{v}(X)(\log d)^{\frac{3}{4}}$ as $\sigma_*(X) \le \tilde{v}(X)$. ☐

6.3. **Proof of Corollary 2.2.** The deduction of Corollary 2.2 from Theorem 2.1 is nearly immediate; we spell out the details for completeness.

Proof of Corollary 2.2. When $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$ are self-adjoint, the probability bound follows immediately from Theorem 2.1. This bound extends directly to general $A_0, \ldots, A_n \in \mathrm{M}_d(\mathbb{C})$ by Remark 2.6. The bound on the expectation is now obtained by integrating the probability bound. More precisely, we have

$$\begin{aligned}
&\mathbf{E}[(\|X\| - \|X_{\mathrm{free}}\| - C\tilde{v}(X)(\log d)^{\frac{3}{4}})_+] \\
&= \int_0^\infty \mathbf{P}[\|X\| \ge \|X_{\mathrm{free}}\| + C\tilde{v}(X)(\log d)^{\frac{3}{4}} + s]\, ds \\
&\le \int_0^\infty e^{-s^2/C^2\sigma_*(X)^2}\, ds = C'\sigma_*(X)
\end{aligned}$$

for a universal constant C'. It follows that

$$\mathbf{E}\|X\| \le \|X_{\mathrm{free}}\| + C\tilde{v}(X)(\log d)^{\frac{3}{4}} + C'\sigma_*(X).$$

It remains to note that as $\sigma_*(X) \le \tilde{v}(X)$, the last term may be eliminated at the expense of choosing a slightly larger universal constant C. ☐

7. Strong asymptotic freeness

The aim of this section is to prove our results on asymptotic freeness that were formulated in Section 2.3. The proof of Theorem 2.10 is divided into two parts. In Section 7.1 we will prove weak asymptotic freeness (part a). This part of the proof is elementary and uses only the basic estimates of Section 4.2; when specialized to Wigner matrices, it yields a self-contained proof of Voiculescu's Theorem 4.3. In Section 7.2, we will prove strong asymptotic freeness (part b) by combining Theorem 2.1 with the linearization trick of [19] and concentration estimates. Finally, Corollary 2.11 will be deduced from Theorem 2.10 in Section 7.3.

7.1. Weak asymptotic freeness.

The aim of this section is to prove part a of Theorem 2.10. By linearity of the trace, it evidently suffices to assume

$$p(H_1,\ldots,H_m) = H_{k_1}\cdots H_{k_q}$$

is a monomial of degree q for some $q\in\mathbb{N}$ and $1\le k_1,\ldots,k_q\le m$. This assumption will be made throughout the proof of part a of Theorem 2.10.

Throughout this section, we let $H_1^N,\ldots,H_m^N$ be defined as in Theorem 2.10. We begin with some preliminary observations. First, we note the following.

Lemma 7.1. *We have* $\sup_{N,k}\mathbf{E}[\mathrm{tr}\,|H_k^N-\mathbf{E}[H_k^N]|^q]^{\frac{1}{q}}<\infty$ *for every* $q\in\mathbb{N}$.

Proof. By assumption, $\sigma(H_k^N)^2 = \|\mathbf{E}(H_k^N)^2\| = 1+o(1)$. The conclusion follows from the noncommutative Khintchine inequality, cf. [28, §9.8] or [39, §3.1]. □

Before we proceed to the main part of the proof, we perform a simple reduction: we show that it suffices to assume $\mathbf{E}[H_k^N]=0$. This elementary observation will avoid unnecessary notational complications.

Lemma 7.2. *Denote* $\bar H_k^N := H_k^N-\mathbf{E}[H_k^N]$. *Then we have*

$$\lim_{N\to\infty}\mathbf{E}\,\mathrm{tr}\,|H_{k_1}^N\cdots H_{k_q}^N-\bar H_{k_1}^N\cdots\bar H_{k_q}^N| = 0.$$

Proof. Note that

$$H_{k_1}^N\cdots H_{k_q}^N-\bar H_{k_1}^N\cdots\bar H_{k_q}^N = \sum_{l=1}^q \bar H_{k_1}^N\cdots\bar H_{k_{l-1}}^N\mathbf{E}[H_{k_l}^N]H_{k_{l+1}}^N\cdots H_{k_q}^N.$$

Thus

$$\mathbf{E}\,\mathrm{tr}\,|H_{k_1}^N\cdots H_{k_q}^N-\bar H_{k_1}^N\cdots\bar H_{k_q}^N| \le q\max_{k,l}\|\mathbf{E}[H_k^N]\|\{(\mathbf{E}\,\mathrm{tr}\,|\bar H_l^N|^q)^{\frac{1}{q}}+\|\mathbf{E}[H_l^N]\|\}^{q-1}$$

by Hölder's inequality. As $\|\mathbf{E}[H_k^N]\|=o(1)$, it remains to note that $\mathbf{E}\,\mathrm{tr}\,|\bar H_k^N|^q$ is uniformly bounded as $N\to\infty$ by Lemma 7.1. □

By Lemma 7.2, we can assume without loss of generality in the remainder of the proof of part a of Theorem 2.10 that $\mathbf{E}[H_k^N]=0$ for all k.

We now turn the the main part of the proof. The basic tool we will use is the classical Wick formula for Gaussian moments [27, Theorem 22.3], which should be compared with its free counterpart in Definition 4.2.

Lemma 7.3 (Wick formula)**.** *Let* $g_1,\ldots,g_n$ *be i.i.d. standard Gaussians. Then*

$$\mathbf{E}[g_{k_1}\cdots g_{k_q}] = \sum_{\pi\in\mathrm{P}_2([q])}\prod_{\{i,j\}\in\pi}\delta_{k_ik_j}$$

for every $q\ge 1$ *and* $k_1,\ldots,k_q\in[n]$.

From the Wick formula, we deduce the following.

Corollary 7.4. *Suppose* $\mathbf{E}[H_k^N]=0$ *for all* $k\in[m]$, *and let* $\mathrm{k}=(k_1,\ldots,k_q)$. *Then*

$$\mathbf{E}[\mathrm{tr}\,H_{k_1}^N\cdots H_{k_q}^N]=\sum_{\pi\in\mathrm{P}_2([q])}\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N]\prod_{\{r,s\}\in\pi}\delta_{k_rk_s},$$

where $H_{1|\pi,\mathrm{k}}^N,\ldots,H_{q|\pi,\mathrm{k}}^N$ *are jointly Gaussian random matrices defined as follows:*

1. $H_{r|\pi,\mathrm{k}}^N$ *has the same distribution as* $H_{k_r}^N$.
2. $H_{r|\pi,\mathrm{k}}^N=H_{s|\pi,\mathrm{k}}^N$ *if* $\{r,s\}\in\pi$.
3. $H_{r|\pi,\mathrm{k}}^N$ *and* $H_{s|\pi,\mathrm{k}}^N$ *are independent if* $r\neq s$, $\{r,s\}\notin\pi$.

Proof. As $\mathbf{E}[H_k^N]=0$, we may write

$$H_k^N=\sum_{i=1}^n g_{ki}\,A_{ki},$$

where g_{ki} are i.i.d. standard Gaussians and $A_{ki}\in\mathrm{M}_d(\mathbb{C})_{\mathrm{sa}}$. Then

$$\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N]\prod_{\{r,s\}\in\pi}\delta_{k_rk_s}=\sum_{i_1,\ldots,i_q}\mathrm{tr}\,A_{k_1i_1}\cdots A_{k_qi_q}\prod_{\{r,s\}\in\pi}\delta_{k_rk_s}\delta_{i_ri_s}$$

by construction. On the other hand

$$\mathbf{E}[\mathrm{tr}\,H_{k_1}^N\cdots H_{k_q}^N]=\sum_{i_1,\ldots,i_q}\mathrm{tr}\,A_{k_1i_1}\cdots A_{k_qi_q}\sum_{\pi\in\mathrm{P}_2([q])}\prod_{\{r,s\}\in\pi}\delta_{k_rk_s}\delta_{i_ri_s}$$

by Lemma 7.3, completing the proof. □

The main idea that gives rise to weak asymptotic freeness is that the terms in Corollary 7.4 that correspond to crossing pairings are asymptotically negligible. This will follow readily from the following lemma.

Lemma 7.5. *In the setting of Corollary 7.4, we have*

$$|\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N]|\le\max_{k,l}w(H_k^N,H_l^N)^4\max_k\mathbf{E}[\mathrm{tr}\,|H_k^N|^{q-4}]$$

for any crossing pairing $\pi\in\mathrm{P}_2([q])\backslash\mathrm{NC}_2([q])$ *such that* $k_r=k_s$ *for all* $\{r,s\}\in\pi$.

Proof. By assumption, the exist $\{r_1,s_1\},\{r_2,s_2\}\in\pi$ such that $r_1<r_2<s_1<s_2$. Computing the expectation with respect to these indices only yields

$$\begin{aligned}&\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N]=\\&\quad\sum_{i,j}\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{r_1-1|\pi,\mathrm{k}}^N A_{k_{r_1}i}H_{r_1+1|\pi,\mathrm{k}}^N\cdots H_{r_2-1|\pi,\mathrm{k}}^N A_{k_{r_2}j}H_{r_2+1|\pi,\mathrm{k}}^N\cdots\\&\qquad\qquad H_{s_1-1|\pi,\mathrm{k}}^N A_{k_{r_1}i}H_{s_1+1|\pi,\mathrm{k}}^N\cdots H_{s_2-1|\pi,\mathrm{k}}^N A_{k_{r_2}j}H_{s_2+1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N],\end{aligned}$$

where we used the notation in the proof of Corollary 7.4. Cyclically permuting the trace, applying Lemma 4.5, and using Hölder's inequality yields

$$|\mathbf{E}[\mathrm{tr}\,H_{1|\pi,\mathrm{k}}^N\cdots H_{q|\pi,\mathrm{k}}^N]|\le w(H_{k_{r_1}}^N,H_{k_{r_2}}^N)^4\prod_{l\in[q]\backslash\{r_1,r_2,s_1,s_2\}}\mathbf{E}[\mathrm{tr}\,|H_{k_l}^N|^{q-4}]^{\frac{1}{q-4}}.$$

The conclusion follows readily. □

On the other hand, the assumption $\|\mathbf{E}[(H_k^N)^2]-\mathbf{1}\|\to 0$ implies the following.

Lemma 7.6. *In the setting of Corollary 7.4, we have*

$$\lim_{N\to\infty} \mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] = 1$$

for any noncrossing pairing $\pi \in \mathrm{NC}_2([q])$ *such that* $k_r = k_s$ *for all* $\{r,s\} \in \pi$.

Proof. Any noncrossing pairing $\pi \in \mathrm{NC}_2([q])$ must contain at least one adjacent pair $\{r, r+1\} \in \pi$. By cyclic permutation of the trace, we may assume $\{q-1, q\} \in \pi$. Computing the expectation with respect to this pair yields

$$\mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] = \mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q-2|\pi,\mathrm{k}} \mathbf{E}[(H^N_{k_q})^2]].$$

In particular, we obtain using Hölder's inequality

$$|\mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] - \mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q-2|\pi,\mathrm{k}}]| \\ \le \|\mathbf{E}[(H^N_{k_q})^2] - \mathbf{1}\| \prod_{k=1}^{q-2} \mathbf{E}[\mathrm{tr}\, |H^N_k|^{q-2}]^{\frac{1}{q-2}}.$$

As $\pi \backslash \{\{q-1, q\}\} \in \mathrm{NC}_2([q-2])$, we may iterate this procedure to obtain

$$|\mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] - 1| \le \frac{q}{2} \max_k \|\mathbf{E}[(H^N_k)^2] - \mathbf{1}\| \max_k \max_{l\le q} \mathbf{E}[\mathrm{tr}\, |H^N_k|^l].$$

The conclusion follows as $\|\mathbf{E}[(H^N_k)^2] - \mathbf{1}\| \to 0$ as $N \to \infty$ by assumption, while $\mathbf{E}[\mathrm{tr}\, |H^N_k|^l]$ is uniformly bounded for all $l \le q$ and $N \ge 1$ by Lemma 7.1. □

The proof of weak asymptotic freeness is now readily completed.

Proof of Theorem 2.10: part a. By Lemma 7.2, we may assume without loss of generality that $\mathbf{E}[H^N_k] = 0$ for all k, N. By Lemma 7.6 and Definition 4.2, we have

$$\lim_{N\to\infty} \sum_{\pi \in \mathrm{NC}_2([q])} \mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] \prod_{\{r,s\}\in\pi} \delta_{k_r k_s} = \tau(s_{k_1} \cdots s_{k_q}).$$

On the other hand, Lemma 7.5 and Proposition 4.6 yield

$$\left| \sum_{\pi \in \mathrm{P}_2([q]) \backslash \mathrm{NC}_2([q])} \mathbf{E}[\mathrm{tr}\, H^N_{1|\pi,\mathrm{k}} \cdots H^N_{q|\pi,\mathrm{k}}] \prod_{\{r,s\}\in\pi} \delta_{k_r k_s} \right| \\ \le |\mathrm{P}_2([q])| \max_k v(H^N_k)^2 \sigma(H^N_k)^2 \max_k \mathbf{E}[\mathrm{tr}\, |H^N_k|^{q-4}].$$

As $\sigma(H^N_k)$ and $\mathbf{E}[\mathrm{tr}\, |H^N_k|^{q-4}]$ are uniformly bounded as $N \to \infty$ by Lemma 7.1, the assumption $v(H^N_k) = o(1)$ implies the right-hand side vanishes as $N \to \infty$. Thus

$$\lim_{N\to\infty} \mathbf{E}[\mathrm{tr}\, H^N_{k_1} \cdots H^N_{k_q}] = \tau(s_{k_1} \cdots s_{k_q})$$

for all $q \in \mathbb{N}$ and $1 \le k_1, \ldots, k_q \le m$ by Corollary 7.4. The conclusion extends immediately to any noncommutative polynomial $p(H^N_1, \ldots H^N_m)$ by linearity. □

7.2. **Strong asymptotic freeness.** The main idea behind the proof of part *b* of Theorem 2.10 is that the behavior of polynomials can be controlled by that of associated random matrices of the form (2.1). We have already encountered a very simple form of such a linearization argument in Lemma 3.14, where it was used to obtain nonasymptotic bounds for sample covariance matrices. As we are presently interested in asymptotics, we can directly invoke the abstract linearization argument of Haagerup and Thorbjørnsen [19, Lemma 1 and pp. 758–760].

Theorem 7.7 (Haagerup-Thorbjørnsen)**.** *Suppose that for every $\varepsilon > 0$, $d' \in \mathbb{N}$, and $A_0, \ldots, A_m \in \mathrm{M}_{d'}(\mathbb{C})_{\mathrm{sa}}$, the following holds almost surely:*

$$\mathrm{sp}(A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes H_k^N) \subseteq \mathrm{sp}(A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes s_k) + [-\varepsilon, \varepsilon]$$

eventually as $N \to \infty$. Then

$$\limsup_{N\to\infty} \|p(H_1^N, \ldots, H_m^N)\| \le \|p(s_1, \ldots, s_m)\| \quad a.s.$$

for every noncommutative polynomial p.

Let again $H_1^N, \ldots, H_m^N$ be defined as in Theorem 2.10. Then we may write

$$H_k^N = B_{k0}^N + \sum_{i=1}^{n_k^N} g_{ki}^N B_{ki}^N,$$

where $n_k^N \in \mathbb{N}$, $B_{ki}^N \in \mathrm{M}_{d(N)}(\mathbb{C})_{\mathrm{sa}}$, and $(g_{ki}^N)_{k\in[m],i\in[n_k^N]}$ are i.i.d. standard Gaussians for each N (we need not specify the joint distribution for different N, but we assume all random matrices have been placed on a single probability space). Let us fix in the following any $d' \in \mathbb{N}$ and $A_0, \ldots, A_m \in \mathrm{M}_{d'}(\mathbb{C})_{\mathrm{sa}}$, and define

$$\Xi^N := A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes H_k^N = A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes B_{k0}^N + \sum_{k=1}^m \sum_{i=1}^{n_k^N} (A_k \otimes B_{ki}^N)\, g_{ki}^N$$

and its free analogue

$$\Xi_{\mathrm{free}}^N := A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes B_{k0}^N + \sum_{k=1}^m \sum_{i=1}^{n_k^N} A_k \otimes B_{ki}^N \otimes s_{ki},$$

where $(s_{ki})_{k,i}$ is a free semicircular family. Then we have the following.

Lemma 7.8. *If $v(H_k^N) = o((\log d(N))^{-\frac{3}{2}})$ as $N \to \infty$ for all k, then*

$$\mathrm{sp}(A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes H_k^N) \subseteq \mathrm{sp}(\Xi_{\mathrm{free}}^N) + [-\varepsilon, \varepsilon]$$

eventually as $N \to \infty$ a.s. for every $\varepsilon > 0$.

Proof. As $H_1^N, \ldots, H_m^N$ are independent, we have

$$\mathrm{Cov}(\Xi^N) = \sum_{k=1}^m \mathrm{Cov}(A_k \otimes H_k^N) = \sum_{k=1}^m \|A_k\|_{\mathrm{HS}}^2 \mathrm{Cov}(H_k^N).$$

As $A_1, \ldots, A_m$ are fixed, it follows that $v(\Xi^N) = \|\mathrm{Cov}(\Xi^N)\|^{\frac{1}{2}} = o((\log d(N))^{-\frac{3}{2}})$. On the other hand, we readily compute

$$\sigma(\Xi^N)^2 = \left\| \sum_{k=1}^m A_k^2 \otimes \mathbf{E}[(H_k^N)^2] \right\|,$$

so $\|\mathbf{E}[(H_k^N)^2] - \mathbf{1}\| = o(1)$ implies that $\sigma(\Xi^N) = O(1)$. Therefore

$$\mathbf{P}\big[\mathrm{sp}(\Xi^N) \subseteq \mathrm{sp}(\Xi_{\mathrm{free}}^N) + \varepsilon_N[-1, 1]\big] \ge 1 - e^{-(\log N)^3}$$

by Theorem 2.1 and $d(N) \ge N$, where

$$\varepsilon_N := C\{\tilde{v}(\Xi^N)(\log d' d(N))^{\frac{3}{4}} + \sigma_*(\Xi^N)(\log d(N))^{\frac{3}{2}}\} = o(1)$$

as $\sigma_*(\Xi^N) \le v(\Xi^N)$. It remains to note that as $\sum_N e^{-(\log N)^3} < \infty$, the conclusion follows from the Borel-Cantelli lemma. ∎

On the other hand, $\|\mathbf{E}[(H_k^N)^2] - \mathbf{1}\| = o(1)$ ensures that the spectrum of $\Xi_{\rm free}^N$ concentrates around that of $A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes s_k$. This is the analogue in the present setting of Lemma 7.6 in the previous section. We first prove a special case.

Lemma 7.9. *In the special case that* $\mathbf{E}[H_k^N] = 0$ *and* $\mathbf{E}[(H_k^N)^2] = \mathbf{1}$ *for all* k,

$$\mathrm{sp}(\Xi_{\rm free}^N) = \mathrm{sp}(A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes s_k).$$

Proof. In the present setting, we may write

$$\Xi_{\rm free}^N = A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes H_{k,\rm free}^N,$$

where

$$H_{k,\rm free}^N = \sum_{i=1}^{n_k^N} B_{ki}^N \otimes s_{ki}$$

satisfies $(\mathrm{tr} \otimes \tau)((H_{k,\rm free}^N)^2) = \sum_i (B_{ki}^N)^2 = \mathbf{1}$. By Definition 4.2, we may compute

$$(\mathrm{tr} \otimes \tau)(H_{k_1,\rm free}^N \cdots H_{k_q,\rm free}^N) = \sum_{\pi \in \mathrm{NC}_2([q])} \sum_{i_1,\ldots,i_q} \mathrm{tr}(B_{k_1 i_1}^N \cdots B_{k_q i_q}^N) \prod_{\{r,s\} \in \pi} \delta_{k_r k_s} \delta_{i_r i_s}.$$

It follows exactly as in the proof of Lemma 7.6 that

$$(\mathrm{tr} \otimes \tau)(H_{k_1,\rm free}^N \cdots H_{k_q,\rm free}^N) = \tau(s_{k_1} \cdots s_{k_q})$$

for all $q \in \mathbb{N}$, $1 \le k_1, \ldots, k_q \le m$, and $N \ge 1$. In particular, it follows that

$$(\mathrm{tr} \otimes \tau)((\Xi_{\rm free}^N)^q) = (\mathrm{tr} \otimes \tau)((A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes s_k)^q)$$

for all $q \in \mathbb{N}$. As $\Xi_{\rm free}^N$ is a bounded operator, the equality of all moments implies that the spectral distributions of $\Xi_{\rm free}^N$ and $A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes s_k$ coincide. Therefore, as $\mathrm{tr} \otimes \tau$ is a faithful state, their spectra coincide as well. $\square$

The general case now follows by a perturbation argument.

Lemma 7.10. *When* $\|\mathbf{E}[H_k^N]\| = o(1)$ *and* $\|\mathbf{E}[(H_k^N)^2] - \mathbf{1}\| = o(1)$ *for all* k,

$$\mathrm{sp}(\Xi_{\rm free}^N) \subseteq \mathrm{sp}(A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes s_k) + [-\varepsilon, \varepsilon]$$

eventually as $N \to \infty$ *for every* $\varepsilon > 0$.

Proof. Define

$$\tilde{\Xi}_{\rm free}^N := A_0 \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes \tilde{H}_{k,\rm free}^N,$$

where

$$\tilde{H}_{k,\rm free}^N = \frac{\sum_{i=1}^{n_k^N} B_{ki}^N \otimes s_{ki} + \big(\|\mathbf{E}[(H_k^N)^2]\|\mathbf{1} - \mathbf{E}[(H_k^N)^2]\big)^{\frac{1}{2}} \otimes \tilde{s}_k}{\|\mathbf{E}[(H_k^N)^2]\|^{\frac{1}{2}}}$$

and $(s_{ki}, \tilde{s}_k)_{k,i}$ is a free semicircular family. As by construction $(\mathrm{id} \otimes \tau)(\tilde{H}_{k,\rm free}^N) = 0$ and $(\mathrm{id} \otimes \tau)((\tilde{H}_{k,\rm free}^N)^2) = \mathbf{1}$, Lemma 7.9 implies that

$$\mathrm{sp}(\tilde{\Xi}_{\rm free}^N) = \mathrm{sp}(A_0 \otimes \mathbf{1} + \textstyle\sum_{k=1}^m A_k \otimes s_k).$$

Next, we estimate

$$\|\Xi_{\rm free}^N - \tilde{\Xi}_{\rm free}^N\| \le \sum_{k=1}^m \|A_k\| \{\|\mathbf{E}[H_k^N]\| + \|H_{k,\rm free}^N - \tilde{H}_{k,\rm free}^N\|\},$$

where $H^N_{k,\mathrm{free}}$ is defined in the proof of Lemma 7.9. Moreover, we have

$$\|H^N_{k,\mathrm{free}} - \tilde H^N_{k,\mathrm{free}}\| \le \left|1 - \frac{1}{\|\mathbf{E}[(H^N_k)^2]\|^{\frac12}}\right| \|H^N_{k,\mathrm{free}}\| + \frac{2\big\|\|\mathbf{E}[(H^N_k)^2]\|\mathbf{1} - \mathbf{E}[(H^N_k)^2]\big\|^{\frac12}}{\|\mathbf{E}[(H^N_k)^2]\|^{\frac12}}$$

using $\|\tilde s_k\| = 2$. Now note that $\|\mathbf{E}[H^N_k]\| = o(1)$ and $\|\mathbf{E}[(H^N_k)^2] - \mathbf{1}\| = o(1)$ imply $\|H^N_{k,\mathrm{free}}\| = O(1)$ by Lemma 2.5. Thus the above expressions yield

$$\lim_{N\to\infty} \|\Xi^N_{\mathrm{free}} - \tilde\Xi^N_{\mathrm{free}}\| = 0.$$

In particular, this implies by (6.2) that

$$\|(z\mathbf{1} - \Xi^N_{\mathrm{free}})^{-1}\| \le \|(z\mathbf{1} - \tilde\Xi^N_{\mathrm{free}})^{-1}\| + \frac{\varepsilon}{(\operatorname{Im} z)^2}$$

for all $z \in \mathbb{C}$, $\operatorname{Im} z > 0$ holds eventually as $N \to \infty$ for every $\varepsilon > 0$. The conclusion now follows by invoking Lemma 6.4. □

Before we can conclude the proof, we require a concentration argument.

Lemma 7.11. *If $v(H^N_k) = o((\log d(N))^{-\frac32})$ as $N \to \infty$ for all k, then*

$$\lim_{N\to\infty} |\|p(H^N_1,\ldots,H^N_m)\| - \mathbf{E}[\|p(H^N_1,\ldots,H^N_m)\|]| = 0 \quad a.s.,$$
$$\lim_{N\to\infty} |\operatorname{tr} p(H^N_1,\ldots,H^N_m) - \mathbf{E}[\operatorname{tr} p(H^N_1,\ldots,H^N_m)]| = 0 \quad a.s.$$

for every noncommutative polynomial p.

Proof. Fix a noncommutative polynomial p of degree q. Define a function f either as $f(g) = \|p(H^N_1,\ldots,H^N_m)\|$ or $f(g) = \operatorname{tr} p(H^N_1,\ldots,H^N_m)$, where $g = (g^N_{ki})_{k\in[m],i\in[n^N_k]}$. We may assume without loss of generality that $n^N_k \le d(N)^2$ as in the proof of Lemma 4.7, so the random vector g has dimension at most $md(N)^2$.

We begin by estimating as in the proofs of Lemma 7.2 and Corollary 4.14 that

$$|f(g) - f(g')| \le L\|g - g'\|, \qquad L = C(p)4^{q-1}\max_k \sigma_*(H^N_k)$$

for all $g, g' \in \Omega$, where

$$\Omega := \{g : \|H^N_k\| \le 4 \text{ for all } k\}$$

and $C(p)$ is a constant that depends only on the polynomial p.

By Corollary 2.2 and a union bound, we can estimate

$$\mathbf{P}[\Omega^c] \le \sum_{k=1}^m \mathbf{P}[\|H^N_k\| > 4] \le me^{-(\log d(N))^3}$$

eventually as $N \to \infty$, where we used that $\sigma_*(H^N_k) \le v(H^N_k) = o((\log d(N))^{-\frac32})$ and $\|H^N_{k,\mathrm{free}}\| \le \|\mathbf{E}[H^N_k]\| + 2\sigma(H^N_k) = 2 + o(1)$ by Lemma 2.5.

As f is L-Lipschitz on Ω, the classical Lipschitz extension theorem of Kirszbraun ensures the existence of a globally L-Lipschitz function $\tilde f$ such that $\tilde f(g) = f(g)$ for $g \in \Omega$. We can therefore estimate for sufficiently large N

$$\begin{aligned} |\mathbf{E}[f(g)] - \mathbf{E}[\tilde f(g)]| &= |\mathbf{E}[(f(g) - \tilde f(g))1_{\Omega^c}]| \\ &\le \mathbf{P}[\Omega^c]^{\frac12}\{(\mathbf{E}|f(g)|^2)^{\frac12} + (\mathbf{E}|\tilde f(g)|^2)^{\frac12}\}. \\ &\le \mathbf{P}[\Omega^c]^{\frac12}\{(\mathbf{E}|f(g)|^2)^{\frac12} + |f(0)| + L\sqrt{m}d(N)\}, \end{aligned}$$

where we used Cauchy-Schwarz and that $0 \in \Omega$ for sufficiently large N. Now note that $(\mathbf{E}|f(g)|^2)^{\frac{1}{2}} \lesssim 1 + \max_k(\mathbf{E}\|H_k^N\|^{2q})^{\frac{1}{2q}}$ by Hölder's inequality, with a universal constant depending on p only. It therefore follows from Corollary 2.2 that $(\mathbf{E}|f(g)|^2)^{\frac{1}{2}}$ is uniformly bounded as $N \to \infty$. As $|f(0)|$ is clearly also uniformly bounded, the estimate $\mathbf{P}[\Omega^c] \le me^{-(\log d(N))^3}$ implies that

$$|\mathbf{E}[f(g)] - \mathbf{E}[\tilde{f}(g)]| = o(1)$$

as $N \to \infty$. On the other hand, we can compute

$$\begin{aligned}\mathbf{P}[|f(g) - \mathbf{E}[\tilde{f}(g)]| \ge L\log N] &\le \mathbf{P}[\Omega^c] + \mathbf{P}[|\tilde{f}(g) - \mathbf{E}[\tilde{f}(g)]| \ge L\log N] \\ &\le me^{-(\log N)^3} + 2e^{-\frac{(\log N)^2}{2}}\end{aligned}$$

by Lemma 4.13 and $d(N) \ge N$. Thus

$$|f(g) - \mathbf{E}[f(g)]| \le L\log N + o(1)$$

eventually as $N \to \infty$ a.s. by the Borel-Cantelli lemma. But as $\sigma_*(H_k^N) \le v(H_k^N) = o((\log N)^{-\frac{3}{2}})$, we have $L\log N = o(1)$ as $N \to \infty$, and the proof is complete. □

We can now complete the proof of Theorem 2.10.

Proof of Theorem 2.10: part b. Theorem 7.7 and Lemmas 7.8 and 7.10 yield

$$\limsup_{N\to\infty} \|p(H_1^N, \ldots, H_m^N)\| \le \|p(s_1, \ldots, s_m)\| \quad \text{a.s.}$$

for every noncommutative polynomial p. On the other hand, combining part a of Theorem 2.10 with Lemma 7.11 yields that

$$\lim_{N\to\infty} \operatorname{tr} p(H_1^N, \ldots, H_m^N) = \tau(p(s_1, \ldots, s_m)) \quad \text{a.s.}$$

The latter implies

$$\liminf_{N\to\infty} \|p(H_1^N, \ldots, H_m^N)\| \ge \liminf_{N\to\infty} \operatorname{tr}(|p(H_1^N, \ldots, H_m^N)|^r)^{\frac{1}{r}} = \tau(|p(s_1, \ldots, s_m)|^r)^{\frac{1}{r}}$$

a.s. for every $r \in \mathbb{N}$, where we used that $|p(H_1^N, \ldots, H_m^N)|^r$ is again a noncommutative polynomial. Letting $r \to \infty$ shows that

$$\lim_{N\to\infty} \|p(H_1^N, \ldots, H_m^N)\| = \|p(s_1, \ldots, s_m)\| \quad \text{a.s.}$$

It remains to note that

$$\lim_{N\to\infty} \mathbf{E}\|p(H_1^N, \ldots, H_m^N)\| = \|p(s_1, \ldots, s_m)\|$$

now follows from Lemma 7.11. □

7.3. **Proof of Corollary 2.11.** We finally deduce Corollary 2.11.

Proof of Corollary 2.11. Applying Theorem 2.10 to $p(H^N) = (H^N)^r$ yields

$$\lim_{N\to\infty} \|H^N\| = \|s\| \quad \text{and} \quad \lim_{N\to\infty} \operatorname{tr}[(H^N)^r] = \tau(s^r) \quad \text{a.s.}$$

for every $r \in \mathbb{N}$, where s is a semicircular variable. As

$$\operatorname{tr}[(H^N)^r] = \int x^r\,\mu_{H^N}(dx), \qquad \tau(s^r) = \int x^r\,\mu_{\mathrm{sc}}(dx),$$

and as μ_{sc} has bounded support, the first conclusion follows as moment convergence implies weak convergence [27, p. 116]. The second conclusion follows as $\|s\| = 2$. □

8. Matrix concentration inequalities

The aim of this section is to prove Theorem 2.12. Throughout this section, we will fix the random matrix model X and define $\sigma_1, \sigma_2, v, L, \sigma_*, R$ as in the statement of Theorem 2.12. We begin with a routine symmetrization argument.

Lemma 8.1. *Let $g_1, \ldots, g_n$ be i.i.d. standard Gaussians independent of X, let $Z_1', \ldots, Z_n'$ be independent copies of $Z_1, \ldots, Z_n$, and define $\tilde{Z}_i := Z_i - Z_i'$. Then*

$$\mathbf{E}\|X\| \le \sqrt{\frac{\pi}{2}}\,\mathbf{E}\Bigg\|\sum_{i=1}^n g_i \tilde{Z}_i\Bigg\|.$$

Proof. Note first that

$$\mathbf{E}\|X\| = \mathbf{E}\Bigg\|\sum_{i=1}^n (Z_i - \mathbf{E}[Z_i'])\Bigg\| \le \mathbf{E}\Bigg\|\sum_{i=1}^n (Z_i - Z_i')\Bigg\|$$

by Jensen's inequality. As (Z_i, Z_i') are exchangeable, the variables $Z_i - Z_i'$ and $\mathrm{sign}(g_i)(Z_i - Z_i')$ have the same distribution. We therefore have

$$\mathbf{E}\|X\| \le \mathbf{E}\Bigg\|\sum_{i=1}^n \mathrm{sign}(g_i)\tilde{Z}_i\Bigg\| \le \sqrt{\frac{\pi}{2}}\,\mathbf{E}\Bigg\|\sum_{i=1}^n g_i \tilde{Z}_i\Bigg\|,$$

where the last step follows by Jensen's inequality and the fact that $\sqrt{\frac{2}{\pi}}\,\mathrm{sign}(g_i) = \mathrm{sign}(g_i)\mathbf{E}|g_i|$ and that $|g_i|$ is independent of $\mathrm{sign}(g_i)$. ☐

The idea is now that if we condition on $\tilde{Z}_1, \ldots, \tilde{Z}_n$, what is left on the right-hand side of Lemma 8.1 is a Gaussian random matrix. This yields the following.

Corollary 8.2. *We have*

$$\mathbf{E}\|X\| \le (1+\varepsilon)\sqrt{\frac{\pi}{2}}\left(\mathbf{E}\Bigg\|\sum_{i=1}^n \tilde{Z}_i^*\tilde{Z}_i\Bigg\|^{\frac{1}{2}} + \mathbf{E}\Bigg\|\sum_{i=1}^n \tilde{Z}_i\tilde{Z}_i^*\Bigg\|^{\frac{1}{2}}\right) + \frac{C}{\varepsilon}\mathbf{E}\left(\sup_{\|M\|_{\mathrm{HS}}\le 1}\sum_{i=1}^n |\mathrm{Tr}[\tilde{Z}_i M]|^2\right)^{\frac{1}{2}} (\log d)^{\frac{3}{2}}$$

for any $\varepsilon > 0$, where C is a universal constant.

Proof. It suffices to estimate the conditional expectation $\mathbf{E}[\|\sum_{i=1}^n g_i\tilde{Z}_i\| | \tilde{Z}_1, \ldots, \tilde{Z}_n]$ using Corollary 2.2, Lemma 2.5, and Young's inequality. ☐

To bound each of the terms that appear in Corollary 8.2, we will use the following variant of the noncommutative Khintchine inequality [36, Theorem 5.1].

Lemma 8.3. *Let $V_1, \ldots, V_n$ be arbitrary independent positive semidefinite $d \times d$ random matrices. Then we have*

$$\mathbf{E}\Bigg\|\sum_{i=1}^n V_i\Bigg\|^{\frac{1}{2}} \le \Bigg\|\sum_{i=1}^n \mathbf{E}[V_i]\Bigg\|^{\frac{1}{2}} + C\left(\mathbf{E}\max_{i\le n}\|V_i\|\right)^{\frac{1}{2}}\sqrt{\log d},$$

where C is a universal constant.

We are now ready to complete the proof of Theorem 2.12.

Proof of Theorem 2.12. An immediate application of Lemma 8.3 yields

$$\mathbf{E}\left\|\sum_{i=1}^n \tilde{Z}_i^*\tilde{Z}_i\right\|^{\frac{1}{2}} \leq \sqrt{2}\,\sigma_1 + CL\sqrt{\log d},$$

where we used $\sum_{i=1}^n \mathbf{E}[\tilde{Z}_i^*\tilde{Z}_i] = 2\,\mathbf{E}[X^*X]$ and $\|\tilde{Z}_i\| \leq \|Z_i\|_{\mathrm{HS}} + \|Z_i'\|_{\mathrm{HS}}$. The analogous bound holds for the second term on the right-hand side of Corollary 8.2 if we replace σ_1 by σ_2. On the other hand, let $\iota : \mathrm{M}_d(\mathbb{C}) \to \mathbb{C}^{d^2}$ map a matrix to its vector of entries. Then we may estimate using Lemma 8.3

$$\mathbf{E}\left(\sup_{\|M\|_{\mathrm{HS}}\leq 1}\sum_{i=1}^n |\mathrm{Tr}[\tilde{Z}_iM]|^2\right)^{\frac{1}{2}} = \mathbf{E}\left\|\sum_{i=1}^n \iota(\tilde{Z}_i)\iota(\tilde{Z}_i)^*\right\|^{\frac{1}{2}} \leq \sqrt{2}\,v + CL\sqrt{\log d},$$

where we used $\sum_{i=1}^n \mathbf{E}[\iota(\tilde{Z}_i)\iota(\tilde{Z}_i)^*] = 2\,\mathrm{Cov}(X)$ and $\|\iota(\tilde{Z}_i)\| \leq \|Z_i\|_{\mathrm{HS}} + \|Z_i'\|_{\mathrm{HS}}$. Combining the above two bounds with Corollary 8.2 yields

$$\mathbf{E}\|X\| \leq (1+\varepsilon)\sqrt{\pi}\{\sigma_1+\sigma_2\} + C(1+\varepsilon)L\sqrt{\log d} + \frac{C}{\varepsilon}(\log d)^{\frac{3}{2}}v + \frac{C}{\varepsilon}(\log d)^2 L$$

for a universal constant C and any $\varepsilon > 0$. As $\sqrt{\pi} < 2$, the first statement of Theorem 2.12 follows by choosing ε sufficiently small.

To deduce the tail bound, we apply Talagrand's concentration inequality for the suprema of empirical processes [23, Corollary 7.9] to the quantity

$$\begin{aligned}\|X\| &= \sup_{\|v\|=\|w\|=1}\left|\sum_{i=1}^n \langle v, Z_iw\rangle\right| \\ &= \sup_{\|v\|=\|w\|=1}\sup_{s\in[0,1]}\sum_{i=1}^n \{s\,\mathrm{Re}\langle v, Z_iw\rangle + \sqrt{1-s^2}\,\mathrm{Im}\langle v, Z_iw\rangle\}.\end{aligned}$$

As $\sup_{\|v\|=\|w\|=1}\sum_{i=1}^n \mathbf{E}[|\langle v, Z_iw\rangle|^2] = \sigma_*^2$, Talagrand's inequality yields

$$\mathbf{P}[\|X\| \geq (1+\varepsilon)\mathbf{E}\|X\| + C\sigma_*\sqrt{t} + C(1+\varepsilon^{-1})Rt\}] \leq e^{-t}$$

for all $t, \varepsilon \geq 0$, where C is a universal constant. Inserting the bound on $\mathbf{E}\|X\|$ obtained above yields the second statement of Theorem 2.12. ☐

9. Discussion and open questions

The aim of this final section is to discuss a number of broader questions that arise from our main results. We first discuss in some detail to what extent the parameter $v(X)$ that quantifies noncommutativity in our bounds is natural, and whether one might hope to improve fundamentally on this parameter. We then proceed to highlight a number of open questions that arise from our results.

9.1. A canonical parameter $\sigma_{**}(X)$ cannot exist.

9.1.1. *Is $v(X)$ a natural parameter?* In all the results of this paper, the presence of noncommutativity and of "intrinsic freeness" was quantified by the parameter $v(X)$. The utility of this parameter is amply demonstrated by the various examples in Section 3: for example, in the independent entry model, $v(X) \asymp \max_{ij} b_{ij}$ recovers precisely the small parameter that controls the previously known behavior (1.4) in this setting, while various models in Section 3.2 illustrate the significance and near-optimality of our bounds in dependent situations.

Nonetheless, it is not difficult to find examples where both $v(X)$, and the slightly improved parameter $\sup_N w(X_1^N)$ discussed in Remark 5.6, fail to capture the correct behavior of Gaussian random matrices. A particularly disconcerting aspect of these parameters is the following. Let X be any random matrix of the form (2.1); then $X\otimes\mathbf{1}$ is again a model of this form, where we tensor on any finite-dimensional identity matrix. Tensoring on an identity clearly has no effect on the spectrum of the matrix: in particular, $\mathrm{sp}(X\otimes\mathbf{1})=\mathrm{sp}(X)$ and $\sigma(X\otimes\mathbf{1})=\sigma(X)$. This invariance fails dramatically, however, for the parameters $v(X)$ and $w(X)$.

Lemma 9.1. *Let $\mathbf{1}_N$ be the identity in $\mathrm{M}_N(\mathbb{C})$. Then for any self-adjoint $d\times d$ random matrix X of the form (2.1), we have*

$$v(X\otimes\mathbf{1}_N)=\sqrt{N}v(X)\quad\textit{for } N\ge 1,$$
$$w(X\otimes\mathbf{1}_N)=\sigma(X)\qquad\textit{for } N\ge d.$$

Proof. We have $\mathrm{Cov}(X\otimes A)=\mathrm{Cov}(X)\otimes\iota(A)\iota(A)^*$ for any deterministic matrix A, where $\iota:\mathrm{M}_d(\mathbb{C})\to\mathbb{C}^{d^2}$ maps a matrix to its vector of entries. Thus $v(X\otimes A)^2=v(X)^2\|A\|_{\mathrm{HS}}^2$, and the first claim follows as $\|\mathbf{1}_N\|_{\mathrm{HS}}=\sqrt{N}$.

To prove the second claim, let $N\ge d$, and define $U\in\mathrm{M}_d(\mathbb{C})\otimes\mathrm{M}_N(\mathbb{C})$ by $U(e_i\otimes e_j)=e_j\otimes e_i$ for $i,j\in[d]$ and $U(e_i\otimes e_j)=0$ otherwise. Then $\|U\|=1$ and

$$\sum_{i,j}(A_i\otimes\mathbf{1})U(A_j\otimes\mathbf{1})U(A_i\otimes\mathbf{1})U(A_j\otimes\mathbf{1})U=\sum_i A_i^2\otimes P\bigg(\sum_i A_i^2\bigg)P^*,$$

where $P:\mathbb{C}^d\to\mathbb{C}^N$ denotes the canoncial embedding $Pe_i=e_i$. Thus $w(X\otimes\mathbf{1})\ge\sigma(X)$ by the last equation display in the proof of Lemma 4.5. On the other hand, $w(X\otimes\mathbf{1})\le\sigma(X\otimes\mathbf{1})=\sigma(X)$ by [37, Proposition 3.2]. □

Lemma 9.1 shows that no matter how well our bounds capture the behavior of the random matrix X, applying our results to $X\otimes\mathbf{1}_d$ can never yield any improvement over the noncommutative Khintchine inequlity (1.2)—despite that tensoring an identity has no effect on the spectrum of the matrix. This observation may lead one to conjecture that the theory of this paper should admit a far-reaching improvement, in which $v(X)$ is replaced by a "natural" parameter that captures correctly the behavior of the spectrum. For example, it was conjectured in [37, 39, 4] that there exist bounds of the kind that are studied in this paper, where the parameter $v(X)$ is replaced by the "natural" parameter $\sigma_*(X)$.

Somewhat surprisingly, such conjectures turn out to be ill-founded. We will presently show that the kind of behavior that is captured by Lemma 9.1 is a fundamental feature of any bound of the form (1.5).

9.1.2. *An impossibility theorem.* Suppose we are given a matrix parameter $\sigma_{**}(X)$ such that the inequality for $d\times d$ centered Gaussian random matrices

$$\mathbf{E}\|X\|\le C\sigma(X)+C\sigma_{**}(X)(\log d)^\beta \tag{9.1}$$

is valid for universal constants $C,\beta>0$. In view of the above discussion, we may aim to find an inequality (9.1) that respects the simplest properties of the spectral norm: the triangle inequlity $\|X+Y\|\le\|X\|+\|Y\|$; unitary invariance $\|U^*XU\|=\|X\|$; and tensor invariance $\|X\otimes\mathbf{1}\|=\|X\|$. Note that all three properties are satisfied also by the parameter $\sigma(X)$. In order for (9.1) to respect these properties, one would have to assume that the parameter $\sigma_{**}(X)$ satisfies these properties up to a universal constant. Let us formalize these requirements as follows:

(1) $\sigma_{**}(X_1 + X_2) \leq C'\{\sigma_{**}(X_1) + \sigma_{**}(X_2)\}$.

(2) $\sigma_{**}(U^* X U) \leq C'\sigma_{**}(X)$ for any non-random unitary matrix U.

(3) $\sigma_{**}(X \otimes \mathbf{1}_N) \leq C'\sigma_{**}(X)$ for any $N \in \mathbb{N}$.

Here C' always denotes a universal constant.

The noncommutative Khintchine inequality (1.2), which corresponds to the case $\sigma_{**}(X) = \sigma(X)$, satisfies all the above requirements but does not capture any noncommutativity. We therefore introduce as a further assumption that the second term of (9.1) becomes negligible at least in the simplest model of random matrix theory, the standard Wigner matrices G^N of Definition 1.1.

(4) $\sigma_{**}(G^N) = o((\log N)^{-\beta})$ as $N \to \infty$.

Remarkably, the above very natural properties prove to be mutually contradictory.

Proposition 9.2. *Suppose that* (9.1) *is valid for some universal constants* C, β. *Then at least one of the properties (1)–(4) must fail for any choice of* C'.

Proof. Let $G_1^N, \ldots, G_n^N$ be i.i.d. standard Wigner matrices of dimension N, and consider the N^n-dimensional Gaussian random matrix

$$X_{n,N} = \sum_{k=1}^{n} \underbrace{\mathbf{1}_N \otimes \cdots \otimes \mathbf{1}_N}_{k-1} \otimes G_k^N \otimes \underbrace{\mathbf{1}_N \otimes \cdots \otimes \mathbf{1}_N}_{n-k}.$$

We will show that if properties (1)–(4) hold for some universal constant $C' \geq 1$, this entails a contradiction. Indeed, properties (1)–(3) yield

$$\begin{aligned}
\sigma_{**}(X_{n,N}) &\overset{(1)}{\leq} \sum_{k=1}^{n} (C')^k \, \sigma_{**}(\mathbf{1}_{N^{k-1}} \otimes G_k^N \otimes \mathbf{1}_{N^{n-k}}) \\
&\overset{(2)}{\leq} \sum_{k=1}^{n} (C')^{k+1} \, \sigma_{**}(G_k^N \otimes \mathbf{1}_{N^{n-1}}) \\
&\overset{(3)}{\leq} \sum_{k=1}^{n} (C')^{k+2} \, \sigma_{**}(G_k^N).
\end{aligned}$$

while we may readily compute $\sigma(X_{n,N}) = \sqrt{n}$. Thus we obtain

$$\limsup_{N\to\infty} \mathbf{E}\|X_{n,N}\| \leq C\sqrt{n}$$

by (9.1) and property (4). As $(\mathbf{E}|\|X_{n,N}\| - \mathbf{E}\|X_{n,N}\||^p)^{\frac{1}{p}} \lesssim \sigma_*(X_{n,N})\sqrt{p} = o(1)$ as $N \to \infty$ by a routine application of Corollary 4.14, we further obtain

$$\limsup_{N\to\infty} (\mathbf{E}\|X_{n,N}\|^p)^{\frac{1}{p}} \leq C\sqrt{n}$$

for any $p \in \mathbb{N}$.

On the other hand, denote by Z_k^N the random variable obtained by drawing one of the eigenvalues of G_k^N uniformly at random. Then we may write

$$\mathbf{E}[\operatorname{tr} X_{n,N}^p] = \mathbf{E}\left[\left(\sum_{k=1}^{n} Z_k^N\right)^p\right].$$

Note that $Z_1^N, \ldots, Z_n^N$ are i.i.d. random variables, whose distributions converge weakly to the standard semicircle distribution as $N \to \infty$ by Theorem 4.3. Thus

$$\limsup_{N\to\infty} (\mathbf{E}\|X_{n,N}\|^{2p})^{\frac{1}{2p}} \geq \limsup_{N\to\infty} \mathbf{E}[\operatorname{tr} X_{n,N}^{2p}]^{\frac{1}{2p}} = \mathbf{E}\left[\left(\sum_{k=1}^{n} Z_k\right)^{2p}\right]^{\frac{1}{2p}},$$

where $Z_1, \ldots, Z_n$ are i.i.d. random variables with the standard semicircle distribution. The classical central limit theorem now implies

$$\mathbf{E}[g^{2p}]^{\frac{1}{2p}} \leq \limsup_{n\to\infty} \limsup_{N\to\infty} \frac{(\mathbf{E}\|X_{n,N}\|^{2p})^{\frac{1}{2p}}}{\sqrt{n}} \leq C$$

for any $p \in \mathbb{N}$, where g is a standard Gaussian variable. But this yields the desired contradiction, as $\mathbf{E}[g^{2p}]^{\frac{1}{2p}} \to \infty$ as $p \to \infty$. □

A special case of Proposition 9.2 disproves the conjecture made in [37, 39, 4]: the parameter $\sigma_*(X)$ satisfies all four properties (1)–(4), and thus an inequality of the form (9.1) with $\sigma_{**}(X) = \sigma_*(X)$ cannot hold.

More generally, Proposition 9.2 shows that no parameter $\sigma_{**}(X)$ can be expected to avoid the kind of "unnatural" behavior that was identified in Lemma 9.1. In fact, the proof of Proposition 9.2 suggests a clear explanation of why this must be the case. The operation of tensoring identities makes it possible to create limiting objects that obey the classical (commutative) notion of independence, as opposed to free independence. However, if properties (1)–(4) hold, then such commutative models can give rise to a small parameter $\sigma_{**}(X)$, so that (9.1) would imply that they behave as their free counterparts up to a universal constant (by Lemma 2.5). These two phenomena stand in contradiction.

9.1.3. *The dimension threshold.* The second identity of Lemma 9.1 shows that our results fail to capture any noncommutative behavior when we tensor a random matrix X by an identity of the same dimension. On the other hand, for standard Wigner matrices G^N, we have $\sigma(G^N \otimes \mathbf{1}_{D(N)}) = 1$ and

$$v(G^N \otimes \mathbf{1}_{D(N)}) \asymp \sqrt{\frac{D(N)}{N}} \ll \sigma(G^N \otimes \mathbf{1}_{D(N)})$$

as soon as $D(N) \ll N$. Thus the case where a random matrix is tensored by an identity of proportional dimension appears as the threshold at which our ability to capture "intrinsic freeness" breaks down.

This phenomenon has an unexpected connection to certain questions in the theory of operator algebras. In the rest of this section, let $G_1^N, \ldots, G_m^N, H_1^N, \ldots, H_m^N$ be independent GUE matrices (that is, self-adjoint $N \times N$ matrices with i.i.d. centered complex Gaussian variables of variance $\frac{1}{N}$ on and above the diagonal). In the recent work [20], it was shown that if strong convergence

$$\lim_{N\to\infty} \|p(G_1^N \otimes \mathbf{1}_N, \ldots, G_m^N \otimes \mathbf{1}_N, \mathbf{1}_N \otimes H_1^N, \ldots, \mathbf{1}_N \otimes H_m^N)\| = \|p(s_1 \otimes \mathbf{1}, \ldots, s_m \otimes \mathbf{1}, \mathbf{1} \otimes s_1, \ldots, \mathbf{1} \otimes s_m)\| \quad \text{a.s.}$$

were to hold for all polynomials p,[2] this would settle a conjecture of Peterson and Thom in the theory of Von Neumann algebras. Using the results of this paper,

[2]Throughout this section $\otimes$ always denotes the minimal tensor product of C^*-algebras.

a slightly weaker fact can be proved. As the following result is only tangentially related to the rest of this paper, we will sketch its proof.

Proposition 9.3. *We have*

$$\lim_{N\to\infty} \|p(G_1^N \otimes \mathbf{1}_{D(N)}, \ldots, G_m^N \otimes \mathbf{1}_{D(N)}, \mathbf{1}_N \otimes H_1^{D(N)}, \ldots, \mathbf{1}_N \otimes H_m^{D(N)})\| = \|p(s_1 \otimes \mathbf{1}, \ldots, s_m \otimes \mathbf{1}, \mathbf{1} \otimes s_1, \ldots, \mathbf{1} \otimes s_m)\| \quad a.s.$$

for every noncommutative polynomial p, provided $D(N) = o\big(\frac{N}{(\log N)^3}\big)$.

While this does not suffice for the purpose of [20], which requires $D(N) = N$, the result was previously known only for $D(N) = o(N^{\frac{1}{3}})$ [10, Theorem 1.2].

Sketch of proof of Proposition 9.3. Fix a dimension $d' \in \mathbb{N}$ and self-adjoint matrices $A_0, \ldots, A_m, B_1, \ldots, B_m \in \mathrm{M}_{d'}(\mathbb{C})_{\mathrm{sa}}$. Define the random matrix

$$X^N = A_0 \otimes \mathbf{1}_N \otimes \mathbf{1}_{D(N)} + \sum_{k=1}^m A_k \otimes G_k^N \otimes \mathbf{1}_{D(N)} + \sum_{k=1}^m B_k \otimes \mathbf{1}_N \otimes H_k^{D(N)}.$$

The assumption on $D(N)$ implies that $v(\sum_{k=1}^m A_k \otimes G_k^N \otimes \mathbf{1}_{D(N)}) = o((\log N)^{-\frac{3}{2}})$. As $(G_k^N)_{k\le m}$ and $(H_k^{D(N)})_{k\le m}$ are independent, we can apply Theorem 2.1 conditionally on $(H_k^{D(N)})_{k\le m}$, Lemma 7.9, and the Borel-Cantelli lemma to show that

$$\mathrm{sp}(X^N) \subseteq \mathrm{sp}(A_0\otimes\mathbf{1}\otimes\mathbf{1}_{D(N)}+\textstyle\sum_{k=1}^m A_k\otimes s_k\otimes\mathbf{1}_{D(N)}+\sum_{k=1}^m B_k\otimes\mathbf{1}\otimes H_k^{D(N)})+[-\varepsilon,\varepsilon]$$

eventually as $N \to \infty$ a.s. for every $\varepsilon > 0$.

On the other hand, let $\mathcal{A}$ be the unital C^*-algebra generated by $\{s_1, \ldots, s_m\}$. Then $\mathrm{M}_{d'}(\mathbb{C}) \otimes \mathcal{A}$ is an exact C^*-algebra, cf. [20, p. 27] and the references therein. Therefore, [19, Theorem 9.1] and [11, Proposition 2.1] imply that

$$\mathrm{sp}(A_0 \otimes \mathbf{1} \otimes \mathbf{1}_{D(N)} + \textstyle\sum_{k=1}^m A_k \otimes s_k \otimes \mathbf{1}_{D(N)} + \sum_{k=1}^m B_k \otimes \mathbf{1} \otimes H_k^{D(N)}) \subseteq \mathrm{sp}(A_0 \otimes \mathbf{1} \otimes \mathbf{1} + \sum_{k=1}^m A_k \otimes s_k \otimes \mathbf{1} + \sum_{k=1}^m B_k \otimes \mathbf{1} \otimes s_k) + [-\varepsilon,\varepsilon]$$

eventually as $N \to \infty$ a.s. for every $\varepsilon > 0$. Linearization as in Theorem 7.7 yields

$$\limsup_{N\to\infty} \|p(G_1^N \otimes \mathbf{1}_{D(N)}, \ldots, G_m^N \otimes \mathbf{1}_{D(N)}, \mathbf{1}_N \otimes H_1^{D(N)}, \ldots, \mathbf{1}_N \otimes H_m^{D(N)})\| \le \|p(s_1 \otimes \mathbf{1}, \ldots, s_m \otimes \mathbf{1}, \mathbf{1} \otimes s_1, \ldots, \mathbf{1} \otimes s_m)\| \quad \text{a.s.}$$

for every noncommutative polynomial p. The reverse inequality follows from weak asymptotic freeness of $(G_k^N)_{k\le m}$ and $(H_k^{D(N)})_{k\le m}$ and concentration of measure as in the analogous part of the proof of Theorem 2.10. ☐

9.2. **Open questions.** We conclude this paper by highlighting some basic open questions that arise from our main results.

9.2.1. *Sharp inequalities.* As was explained in the previous section, there cannot exist a canonical inequality of the form (1.5) that captures correctly the structure of all Gaussian random matrices. However, even if we restrict attention to parameters such as $v(X)$, the main results of this paper are slightly suboptimal in the second-order term. For example, Corollary 2.2 does not fully subsume the known sharp results for the independent entry model, as the logarithmic term $(\log d)^{\frac{3}{2}}$ in (3.3) is slightly worse than the term $\sqrt{\log d}$ in (3.2).

The suboptimal power on the logarithm is relevant only for models that are right at the threshold where "intrinsic freeness" breaks down, and is insignificant in most applications. It is nonetheless an interesting question whether the results of this paper can be refined so that they recover the sharp result for the independent entry model. This would be the case, for example, if one could prove that

$$\mathbf{E}\|X\| \overset{?}{\leq} \|X_{\mathrm{free}}\| + Cv(X)\sqrt{\log d}.$$

Corollary 2.2 falls short of such a bound in two ways: it has a suboptimal power on the logarithm $(\log d)^{\frac{3}{4}}$, and it involves the parameter $\tilde{v}(X)$ rather than $v(X)$. (Replacing $\tilde{v}(X)$ by $\sup_N w(X_1^N)$, as in Remark 5.6, would not suffice to recover the sharp behavior of the independent entry model, cf. [37, §3.8].)

9.2.2. *Universality.* The strongest results of this paper apply to Gaussian random matrices, and we made heavy use of Gaussian analysis in our proofs. While we were able to deduce a much more general matrix concentration inequality by symmetrization in Theorem 2.12, this came at the expense of the loss of a universal constant. Such an approach can therefore not be used, for example, to establish strong asymptotic freeness of non-Gaussian random matrices.

It is a question of considerable interest whether the main results of this paper can be extended to non-Gaussian situations with a sharp leading term. For example, in the setting of Theorem 2.12 with $\|Z_i\|_{\mathrm{HS}} \leq M$ a.s. and $\sigma = \sigma_1 \vee \sigma_2$, one might conjecture a sharp matrix concentration inequality of the form

$$\mathbf{E}\|X\| \overset{?}{\leq} \|X_{\mathrm{free}}\| + C(\log d)^{\frac{3}{4}}\sigma^{\frac{1}{2}}v^{\frac{1}{2}} + C(\log d)^2 M,$$

where $X_{\mathrm{free}} \in \mathrm{M}_d(\mathcal{A})$ is now defined such that the real and imaginary parts of its entries form a semicircular family in the sense of [27, Definition 8.15] with the same covariance as the real and imaginary parts of the entries of X.

An analogous extension of Theorem 2.1 would give rise to strong asymptotic freeness of a very large family of non-Gaussian random matrix models. Such results are of interest already in special cases: for example, an analogue of Theorem 2.1 in the setting where the Gaussian variables g_i in (2.1) are replaced by i.i.d. random signs ε_i would establish strong asymptotic freeness of sparse random sign matrices as in Example 3.5. Such an example would show that $O(d\log^4 d)$ random bits in dimension d already suffice to generate strong asymptotic freeness.

The phenomenon that large random matrices behave as their Gaussian counterparts, known as *universality*, has been deeply investigated in classical random matrix theory in the past decade [34]. The above considerations motivate the importance of such questions to the study of nonhomogeneous random matrices.

9.2.3. *Reverse bounds on the spectrum.* The results of Section 2.2 yield two-sided bounds on the spectral statistics of X in terms of X_{free}. In contrast, Section 2.1 only yields one-sided bounds on the support of the spectrum: we show that $\mathrm{sp}(X) \subseteq \mathrm{sp}(X_{\mathrm{free}}) + [-\varepsilon, \varepsilon]$ with high probability. When one is interested in asymptotics, the latter is usually the difficult direction, while the reverse inclusion follows rather easily from weak bounds on the spectral statistics. It is not clear, however, how to obtain nonasymptotic bounds of the form $\mathrm{sp}(X_{\mathrm{free}}) \subseteq \mathrm{sp}(X) + [-\varepsilon, \varepsilon]$.

To illustrate where the difficulty lies, let us derive a two-sided bound on the spectral norm $\|X\|$ from Theorem 2.7. As X is a $d \times d$ matrix, we have

$$d^{-\frac{1}{2p}}\|X\| \leq (\mathrm{tr}\,|X|^{2p})^{\frac{1}{2p}} \leq \|X\|$$

pointwise. Thus Theorem 2.7 and Corollary 4.14 yield

$$\mathbf{E}\|X\| = (1+o(1))\,(\mathrm{tr}\otimes\tau)(|X_{\mathrm{free}}|^{2p})^{\frac{1}{2p}} \quad \text{when} \quad v(X) \ll p^{-\frac{3}{2}} \ll (\log d)^{-\frac{3}{2}}.$$

However, while $(\mathrm{tr}\otimes\tau)(|X_{\mathrm{free}}|^{2p})^{\frac{1}{2p}} \le \|X_{\mathrm{free}}\|$ exactly as for X, it is not clear how to obtain an analogous lower bound. Resolving this issue would require a quantitative understanding of the concentration of the mass of the spectral distribution of X_{free} near the maximum of its support. The results of [1] provide a detailed qualitative picture of the spectral distribution of X_{free} near the edges of its support, but nonasymptotic estimates do not appear to be known. Precisely the same issue arises in the proof of Theorem 2.1: obtaining a reverse bound would require a lower bound on the moments of the resolvent of X_{free} (cf. Lemma 6.5).

Acknowledgments. M.T.B. was supported in part by NSF grant DMS-1856221. R.v.H. was supported in part by NSF grants DMS-1811735 and DMS-2054565, and by the Simons Collaboration on Algorithms & Geometry. The authors thank Tatiana Brailovskaya, Antti Knowles, Gilles Pisier, Dominik Schröder, Joel Tropp, and Nikita Zhivotovskiy for interesting discussions on the topic of this paper.

References

[1] J. Alt, L. Erdős, and T. Krüger. The Dyson equation with linear self-energy: spectral bands, edges and cusps. *Doc. Math.*, 25:1421–1539, 2020.
[2] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[3] B. Au. Traffic distributions of random band matrices. *Electron. J. Probab.*, 23:Paper No. 77, 48, 2018.
[4] A. S. Bandeira. Ten lectures and forty-two open problems in the mathematics of data science, 2015. Lecture notes.
[5] A. S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016.
[6] C. Bordenave and B. Collins. Strong asymptotic freenes for independent uniform variables on compact groups associated to non-trivial representations. 2020. Preprint arxiv:2012.08759.
[7] A. Bose. *Patterned random matrices*. CRC Press, Boca Raton, FL, 2018.
[8] T. T. Cai, R. Han, and A. R. Zhang. On the non-asymptotic concentration of heteroskedastic Wishart-type matrix. 2020. Preprint arxiv:2008.12434.
[9] A.-P. Calderón. Intermediate spaces and interpolation, the complex method. *Studia Math.*, 24:113–190, 1964.
[10] B. Collins, A. Guionnet, and F. Parraud. On the operator norm of non-commutative polynomials in deterministic matrices and iid GUE matrices. 2019. Preprint arxiv:1912.04588.
[11] B. Collins and C. Male. The strong asymptotic freeness of Haar and deterministic matrices. *Ann. Sci. Éc. Norm. Supér. (4)*, 47(1):147–163, 2014.
[12] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 317–366. North-Holland, Amsterdam, 2001.
[13] L. Erdős and P. Mühlbacher. Bounds on the norm of Wigner-type random matrices. *Random Matrices Theory Appl.*, 8(3):1950009, 28, 2019.
[14] M. Fleermann, W. Kirsch, and T. Kriecherbauer. The almost sure semicircle law for random band matrices with dependent entries. *Stochastic Process. Appl.*, 131:172–200, 2021.
[15] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.
[16] F. R. Gantmacher. *The theory of matrices. Vol. 2*. Chelsea Publishing Co., New York, 1959. Translated by K. A. Hirsch.
[17] U. Haagerup, H. Schultz, and S. Thorbjørnsen. A random matrix approach to the lack of projections in $C^*_{\mathrm{red}}(\mathbb{F}_2)$. *Adv. Math.*, 204(1):1–83, 2006.

[18] U. Haagerup and S. Thorbjørnsen. Random matrices and K-theory for exact C^*-algebras. *Doc. Math.*, 4:341–450, 1999.
[19] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math. (2)*, 162(2):711–775, 2005.
[20] B. Hayes. A random matrix approach to the Peterson-Thom conjecture. 2020. Preprint arXiv:2008.12287.
[21] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017.
[22] R. Latała, R. van Handel, and P. Youssef. The dimension-free structure of nonhomogeneous random matrices. *Invent. Math.*, 214(3):1031–1080, 2018.
[23] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
[24] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
[25] F. Lehner. Computing norms of free operators with matrix coefficients. *Amer. J. Math.*, 121(3):453–486, 1999.
[26] J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.
[27] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
[28] G. Pisier. *Introduction to operator space theory*, volume 294 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2003.
[29] G. Pisier and Q. Xu. Non-commutative L^p-spaces. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1459–1517. North-Holland, Amsterdam, 2003.
[30] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.
[31] D. Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *Internat. Math. Res. Notices*, (20):1013–1025, 1996.
[32] M. Talagrand. *Mean field models for spin glasses. Volume I*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, 2011. Basic examples.
[33] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
[34] T. Tao and V. Vu. Random matrices: the universality phenomenon for Wigner ensembles. In *Modern aspects of random matrix theory*, volume 72 of *Proc. Sympos. Appl. Math.*, pages 121–172. Amer. Math. Soc., Providence, RI, 2014.
[35] J. A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8:1–230, 2015.
[36] J. A. Tropp. The expected norm of a sum of independent random matrices: an elementary approach. In *High dimensional probability VII*, volume 71 of *Progr. Probab.*, pages 173–202. Springer, [Cham], 2016.
[37] J. A. Tropp. Second-order matrix concentration inequalities. *Appl. Comput. Harmon. Anal.*, 44(3):700–736, 2018.
[38] R. van Handel. On the spectral norm of Gaussian random matrices. *Trans. Amer. Math. Soc.*, 369(11):8161–8178, 2017.
[39] R. van Handel. Structured random matrices. In *Convexity and concentration*, volume 161 of *IMA Vol. Math. Appl.*, pages 107–156. Springer, New York, 2017.
[40] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

Department of Mathematics, ETH Zürich, Switzerland
Email address: bandeira@math.ethz.ch

Department of Mathematics, University of California Irvine, USA
Email address: marchb@uci.edu

Fine Hall 207, Princeton University, Princeton, NJ 08544, USA
Email address: rvan@math.princeton.edu