RANDOM MATRICES WITH LOG-RANGE CORRELATIONS, AND LOG-SOBOLEV INEQUALITIES

TODD KEMP AND DAVID ZIMMERMANN

ABSTRACT. Let X_N be a symmetric $N \times N$ random matrix whose $\sqrt{N}$-scaled entries are uniformly square integrable. We prove that if the entries of X_N can be partitioned into independent subsets each of size $o(\log N)$, then the empirical eigenvalue distribution of X_N converges weakly to its mean in probability. If the entries are bounded, the convergence is almost sure; if the entries are Gaussian, we prove almost sure convergence with larger blocks of size $o(N^2/\log N)$. This significantly extends the best previously known results on convergence of eigenvalues for matrices with correlated entries, where the partition subsets are blocks and of size $O(1)$. We also prove the strongest known convergence results for eigenvalues of band matrices.

We prove these results developing a new log-Sobolev inequality, generalizing the first author's introduction of mollified log-Sobolev inequalities: we show that if $\mathbf{Y}$ is a bounded random vector and $\mathbf{Z}$ is a standard normal random vector independent from $\mathbf{Y}$, then the law of $\mathbf{Y} + t\mathbf{Z}$ satisfies a log-Sobolev inequality for all $t > 0$, and we give bounds on the optimal log-Sobolev constant.

CONTENTS

1. Introduction 1
2. Concentration Results for Ensembles with Correlated Entries 5
2.1. Guionnet's Approach to Wigner's Law 5
2.2. The Proof of Theorem 1.1 6
2.3. Theorem 1.2, a Generalization, and Applications 9
2.4. Theorem 1.4 and Generalizations 11
3. Mollified Log-Sobolev Inequalities on $\mathbb{R}^d$ 11
3.1. The Proof of Theorem 3.1 12
Acknowledgments 16
References 16

1. INTRODUCTION

Random matrix theory is primarily interested the convergence of statistics associated to the eigenvalues (or singular values) of $N \times N$ matrices whose entries are random variables with a prescribed joint distribution. The field was begun by Wigner in [37, 38], in which he studied the mean bulk behavior of the eigenvalues of what is now called a *Gaussian Orthogonal Ensemble* GOE_N. This is the Gaussian case of a more general class of random matrices now called *Wigner ensembles*: symmetric random matrixes X_N such that the entries of $\sqrt{N}X_N$ are i.i.d. random variables (modulo the symmetry constraint) with sufficiently many finite moments. There are also corresponding complex Hermitian ensembles, non-symmetric / non-Hermitian ensembles, as well as a parallel world of matrices generalizing the GOE_N, defined not via the distribution of entries but rather by invariance properties of the joint distribution. In this paper, we take Wigner ensembles as the starting point.

Kemp supported in part by NSF CAREER Award DMS-1254807.

Given a symmetric matrix X_N, enumerate its eigenvalues $\lambda_1^N \le \cdots \le \lambda_N^N$ in nondecreasing order. The *empirical spectral distribution (ESD)* of X_N is the random point measure

$$\mu_N = \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j^N}. \tag{1.1}$$

Integrating μ_N against the indicator function $\mathbb{1}_B$ yields the random variable counting the number of eigenvalues in B (building up the histogram of the eigenvalues of X_N). In general, the random variables $\int f\, d\mu_N$ for test functions $f\colon \mathbb{R}\to\mathbb{R}$ are called *linear statistics* of the eigenvalues. Wigner's original papers [37, 38] showed that, for the GOE_N, the ESD μ_N converges weakly in expectation to what is now called Wigner's semicircle law: $\sigma(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}\, dx$. To be precise: this means that $\mathbb{E}(\int f\, d\mu_N) \to \int f\, d\sigma$ for each $f \in C_b(\mathbb{R})$. This convergence was later upgraded to weak a.s. convergence. Many more results are known about the fluctuations of μ_N, the spacing between eigenvalues, and the distribution and fluctuations of the largest eigenvalue. The reader may consult the book [1] and its extensive bibliography for more on these endeavors.

There is also a vast literature on *band matrices*. Originally referring to random matrices with entries that are 0 along many diagonals, these are a wider class of random matrix ensembles generalizing Wigner ensembles, where the upper-triangular entries are still independent, but need not be identically distributed (so long as they satisfy some form of uniform regularity). There is a vast literature on band matrices; see, for example, the expansive paper [2] which uses combinatorial and probabilistic methods to establish that a large class of band matrices have ESD converging a.s. to the semicircle law, with Gaussian fluctuations of a similar form to Wigner matrices. (Our Theorem 1.4 below improves on the main result in [2].)

There are comparatively few papers, however, dealing with random matrices with *correlated entries*. In [34], Shlyakhtenko realized that the tools of operator-valued free probability could be used to compute the limit in expectation of certain kinds of block matrices: ensembles X_{kN} possessed of $k\times k$ blocks that have a fixed covariance structure (uniform among the blocks), where the N^2 blocks are independent up to symmetry. The recent papers [10, 11, 3] showed how to explicitly compute the limit ESD for a wide class of such block matrices with Gaussian entries, and used these results to give applications to quantum information theory. Additionally, in [33], a class of these block matrices was studied and proved to converge almost surely, with applications given to signal processing. (The actual ensembles studied in [33, 34, 10, 11, 3] are presented in a different form, with an overall $k\times k$ block structure with $N\times N$ blocks all whose entries are independent; this is just an orthonormal basis change from the description above, and so has the same ESD.) Note that in these block matrices, the limiting ESD is typically not semicircular. The combinatorial methods used to analyze such ensembles do not easily extend beyond the case that k is fixed as $N\to\infty$.

Our main results, Theorems 1.1 and 1.2, give a significant generalization of ESD convergence for block-type matrices, both in terms of allowing k to grow with N, and softening the rigid structure of the partition into independent blocks.

Theorem 1.1. *Let X_N be an $N\times N$ random matrix. Assume that the entries of X_N satisfy the following conditions.*

(1) *The family $\{N[X_N]_{ij}^2\}_{N\in\mathbb{N},1\le i,j\le N}$ is uniformly integrable.*
(2) *For each N, there is a set partition Π_N of $\{(i,j)\colon 1\le i\le j\le N\}$ and a constant $d_N = o(\log N)$ such that each block of Π_N has size $\le d_N$, and the entries $[X_N]_{ij}$ and $[X_N]_{k\ell}$ are independent if (i,j) and (k,ℓ) are not in the same block of Π_N.*

Then the empirical spectral distribution μ_N of X_N converges weakly in probability to its mean:

$$\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right) \to_{\mathbb{P}} 0, \qquad \textit{for all } f\in \mathrm{Lip}(\mathbb{R}). \tag{1.2}$$

If we further assume that the family $\{\sqrt{N}|[X_N]_{ij}|\}_{N\in\mathbb{N},1\le i,j\le N}$ is uniformly bounded, then the convergence in (1.2) *is almost sure.*

We use similar techniques to those used in the proof of Theorem 1.1 to prove the following stronger result in the case of Gaussian entries: under the appropriate uniform integrability conditions, the convergence of the ESD is almost sure, and guaranteed for blocks of much larger size.

Theorem 1.2. *Let X_N be an $N \times N$ random matrix ensemble whose entries are jointly Gaussian. Assume the entries of X_N satisfy the following conditions.*

(1) *The second moments $\{N\mathbb{E}([X_N]_{ij}^2)\}_{N\in\mathbb{N},1\le i,j\le N}$ are uniformly bounded.*
(2) *For each N, there is a set partition Π_N of $\{(i,j)\colon 1 \le i \le j \le N\}$ and a constant $d_N = o(N^2/\log N)$ such that each block of Π_N has size $\le d_N$, and the entries $[X_N]_{ij}$ and $[X_N]_{k\ell}$ are independent if (i,j) and (k,ℓ) are not in the same block of Π_N.*

Then the empirical spectral distribution μ_N of X_N converges weakly almost surely to its mean:

$$\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right) \to 0 \textit{ a.s.} \qquad \textit{for all } \; f \in \mathrm{Lip}(\mathbb{R}). \tag{1.3}$$

Condition (1) in Theorems 1.1 and 1.2 is analogous to the requirement that the second moments of the entries of $\sqrt{N}X_N$ are normalized in Wigner ensembles. Condition (2) generalizes the independent block structure mentioned above; for example, in the ensembles treated in [10, 11, 3] but with k allowed to grow with N (with $k = o(N^2/\log N)$), one gets convergence of the ESD weakly almost surely. In particular, Theorem 1.2 extends the results of those papers even in the case $k = O(1)$, since only convergence in expectation was known before.

Remark 1.3. Note that the conclusion of Theorems 1.1 and 1.2 is that the ESDs of these ensembles concentrate around their means; it is not true that all these ensembles converge in expectation. Rather, our results are that any of these ensembles that *do* converge in expectation *also* converge in probability, or almost surely, as the case may be. In Section 2.3, we discuss some examples where these results can be applied.

While we are most interested in ensembles with correlated entries, one of the main achievements of our method is an improvement on the (first half of the) main result in [2].

Theorem 1.4. *Let $\{\xi_{ij}\colon 1 \le i \le j\}$ be zero mean unit variance i.i.d. random variables. Let $g\colon [0,1]^2 \to \mathbb{R}_+$ be a symmetric, continuous function. If $[X_N]_{ij} = [X_N]_{ji} = N^{-1/2}g(i/N, j/N)^{1/2}\xi_{ij}$, then the empirical spectral distribution of X_N converges weakly in probability to a probability measure on $\mathbb{R}$. (The limit ESD is the semicircle law if $\int_0^1 g(x,y)\,dy = 1$ for each $x \in [0,1]$.) Moreover, if the ξ_{ij} are bounded random variables, or if the common law of the entries ξ_{ij} satisfies a log-Sobolev inequality (cf.* (1.4) *below), then the convergence is almost sure.*

The ensembles addressed in Theorem 1.4 are the typical formulation of *band matrices*, although that name only really applies when the function g has the form $g(x,y) = \mathbb{1}_{|x-y|\le\delta}$ for some $\delta \in (0,1)$. (In order to satisfy the stochasticity condition to get the semicircle law in the limit, one must use *periodic* band matrices, where g is the indicator of the strip $|x-y| \le \delta$ on all of $\mathbb{R}^2$, projected into $[0,1]^2$ via the equivalence relation identifying two points if they differ by an element of $\mathbb{Z}^2$. See [17, 18].) The central theorem in [2] is a proof of (the semicircular case of) Theorem 1.4, assuming that the common law of the entries ξ_{ij} satisfies a *Poincaré inequality* (cf. 3.1 below). Our Theorem 1.4 yields the convergence in complete generality, only assuming finite second moments; moreover, a technical condition on the laws of the entries (similar to the assumption of a Poincaré inequality) yields almost sure convergence.

Remark 1.5. It should be noted that this is only *half* of the main result in [2], where the authors also show that the fluctuations of these ensembles are Gaussian with an explicit covariance determined by the function g. Their methods are largely combinatorial, while ours are analytic/probabilistic.

Theorems 1.1-1.4 are proved below in Section 2. (In fact, in Section 2.3, we prove the more general Theorem 2.11 of which Theorem 1.2 is a special case.) We prove these results using concentration of measure mediated by a powerful coercive inequality: the *log-Sobolev inequality*. A probability measure μ on $\mathbb{R}^d$ satisfies a log-Sobolev inequality with constant c if

$$\operatorname{Ent}_\mu(f^2) \le c \int |\nabla f|^2 \, d\mu \tag{1.4}$$

for all sufficiently integrable positive functions f with $\int f^2 \, d\mu = 1$; here $\operatorname{Ent}_\mu(g) = \int g \log g \, d\mu$ for a μ-probability density g. The inequality (1.4) first appeared in [35] (in a slightly different form, written in terms of $g = f^2$, where the Dirichlet form on the right-hand-side becomes the relative Fisher information of g), in the context of Gaussian measures. It was later rediscovered by Gross [24] who named it a *log-Sobolev inequality*, and used it to prove an important result in constructive quantum field theory. Over the past four decades, it has played an important role probability theory, functional analysis, and differential geometry; see, for example, [5, 6, 14, 16, 20, 21, 22, 23, 27, 30, 31, 32, 36, 40, 41, 42]. There is a big industry of literature devoted to necessary and sufficient conditions for a log-Sobolev inequality to hold; cf. [8, 9, 15, 26, 29].

Many of the above applications rely on uniform concentration of measure bounds that hold for measures satisfying a log-Sobolev inequality; one nice form of these concentration inequalities is called a Herbst inequality, cf. [26], which yields Gaussian concentration of Lipschitz functionals about their mean. Using the Herbst inequality, Guionnet [25] gave a fundamentally new proof of Wigner's semicircle law; this proof automatically generalized to non-Gaussian ensembles whose entries satisfy a log-Sobolev inequality. Motivated in part by this, the second author of the present paper developed a new approximation scheme, the *mollified log-Sobolev inequality*, in [43]: if Y is any bounded random variable and Z is a standard normal random variable independent from Y, then the law of $Y + tZ$ satisfies a log-Sobolev inequality for all $t > 0$, with a constant $c(t)$ that is bounded in terms of an exponential of $\|Y\|_\infty^2/t$. This fact, together with a standard cutoff argument, allowed the second author to generalize Guionnet's technique to give a fully general proof of Wigner's law for all Wigner ensembles.

Independence played a key role in this analysis, due to the fact that log-Sobolev inequalities behave well under products of measures; cf. Lemma 2.2 below. In the setting of current interest, where we no longer have independence, we will need a multivariate version of the mollified log-Sobolev inequality, with sufficient growth bounds on the constant. That is our second main theorem, which is of independent interest.

Theorem 1.6. *Let $\mathbf{Y}$ be a bounded random vector in $\mathbb{R}^d$, and let $\mathbf{Z}$ be a standard centered normal random vector in $\mathbb{R}^d$ (i.e. $\operatorname{Law}_{\mathbf{Z}}(d\mathbf{x}) = (2\pi)^{-d/2} e^{-|\mathbf{x}|^2/2} \, d\mathbf{x}$) independent from $\mathbf{Y}$. For $0 < t \le |||\mathbf{Y}|||_\infty^2$, the measure $\operatorname{Law}_{\mathbf{Y}+t\mathbf{Z}}$ satisfies a log-Sobolev inequality, with constant $c(t)$ satisfying*

$$c(t) \le \left(K_1 d + K_2 \frac{|||\mathbf{Y}|||_\infty^2}{t} \right) |||\mathbf{Y}|||_\infty^2 \exp\left(\frac{4|||\mathbf{Y}|||_\infty^2}{t} \right) \tag{1.5}$$

for some universal constants $K_1, K_2 > 0$.

Theorem 1.6 has a slightly complicated history. An early version of the present paper proved a weaker estimate, that depended exponentially on d. In response, Bardet, Gozlan, Malrieu, and Zitt [12], building on our techniques, sharpened the inequality to the form (1.5), depending only linearly on d. Our proof uses the Lyapunov approach, and relies on an estimate for the best constant in the Poincaré inequality, which we were only able to prove with a dimension-dependent bound in the previous version of this paper. The main contribution to this problem in [12] was a dimension-independent bound on the Poincaré constant. Below, we cite [12, Theorem 1.2] for the Poincaré inequality bound, and proceed with our original proof of the bound on the LSI (which is cited and used in the proof of (1.5) as given in [12, Theorem 1.3]).

Remark 1.7. To further expound on the history of Theorem 1.6: following the second author's paper [43], in [39] the authors generalized mollified log-Sobolev inequalites to $\mathbb{R}^d$ (and with a class of measures more general than compactly-supported), using a version of the Lyapunov approach as we do. However, they gave no quantitative bounds on the log-Sobolev constant, which is crucial to our present analysis.

Remark 1.8. We do not know if the optimal constant grows with dimension. In [12], some evidence is given to support the conjecture that the optimal constant is independent of dimension. For our present purposes, a dimension independent bound of this form would not improve our result in Theorem 1.1. It is the exponential dependence of the constant on $|||\mathbf{Y}|||_\infty$ that forces the blocks to be of size $o(\log N)$; and this dependence is sharp, as shown below in Example **??**.

The remainder of this paper is organized as follows. In Section 2.1, we discuss how the log-Sobolev inequality can be used to yield concentration results for eigenvalues of random matrices. Following this, Section 2.2 gives the proof of Theorem 1.1. Then Section 2.3 proves Theorem 1.2, and a generalization (Theorem 2.11) which allows more general entries than Gaussians, and applies these results to several random matrix models from the literature. Section 2.4 then proves Theorem 1.4 as a corollary to Theorems 1.1 and 2.11, and discusses a generalization of band matrices where these results still apply. Finally, Section 3 is devoted to the proof of Theorem 1.6.

2. Concentration Results for Ensembles with Correlated Entries

2.1. **Guionnet's Approach to Wigner's Law.** Let us fix notation as in the introduction: let X_N be a symmetric random $N \times N$ matrix ensemble with eigenvalues $\lambda_1^N \le \cdots \le \lambda_N^N$, and let μ_N denote the empirical spectral distribution (ESD) of X_N; cf. (1.1). Wigner's law [37, 38] states that μ_N converges weakly a.s. to the semicircle law σ, in the case that X_N is a GOE_N. Wigner's proof proceeded by the method of moments and is fundamentally combinatorial. Analytic approaches (involving fixed point equations, complex PDEs, and orthogonal polynomials) developed over the ensuing decades. An argument based on concentration of measure was provided by Guionnet in [25, p.70, Thm. 6.6]. The result can be stated thus.

Theorem 2.1. (Guionnet). *Let X_N be a symmetric random matrix. If the joint law of entries of $\sqrt{N}X_N$ satisfies a log-Sobolev inequality with constant c, then for all $\epsilon > 0$ and all Lipschitz $f : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{P}\left(\left|\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right)\right| \ge \epsilon\right) \le 2\exp\left(-\frac{N^2\epsilon^2}{c||f||_{\mathrm{Lip}}^2}\right).$$

In fact, in the Wigner ensembe setting, the i.i.d. condition means we really need only assume that the law of *each entry* satisfies a log-Sobolev inequality. This is due to the following result often called *Segal's lemma*; for a proof, see [24, p. 1074, Rk. 3.3].

Lemma 2.2 (Segal's Lemma)**.** *Let ν_1, ν_2 be probability measures on $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, satisfying log-Sobolev inequalities with constants c_1, c_2, respectively. Then the product measure measure $\nu_1 \otimes \nu_2$ on $\mathbb{R}^{d_1+d_2}$ satisfies a log-Sobolev inequality with constant* $\max\{c_1, c_2\}$.

Theorem 2.1 explicitly gives weak convergence in probability of μ_N to its limit mean. Moreover, in the Wigner ensemble case where the constant c is determined by the common law of the entries and so doesn't depend on N, the rate of convergence is fast enough that a standard Borel–Cantelli argument immediately upgrades this to a.s. convergence. In [43], the second author showed that, under certain integrability conditions, the empirical law of eigenvalues μ_N converges weakly in probability to its mean, *regardless* of whether or not the joint laws of entries satisfy a log-Sobolev inequality. The idea is to use the mollified log-Sobolev inequality (the $d = 1$ case of Theorem 1.6) applied to a cutoff of X_N with GOE_N noise added in with variance t, and then let $t \downarrow 0$.

For our present purposes, where we no longer assume independence or identical distribution of the entries of X_N, it will not suffice to assume each entry satisfies a (mollified) log-Sobolev inequality, which is why we state Guionnet's result as such in Theorem 2.1. Guionnet proved the theorem from the Herbst concentration inequality [26], which shows that Lipschitz functionals of a random variable whose law satisfies a log-Sobolev inequality have sub-Gaussian tails (with dimension-independent bounds determined by the Lipschitz norm of the functional). Theorem 2.1 is then proved by combining this with Lipschitz functional calculus, together with the following lemma from matrix theory (see [28, p.37, Thm. 1, and p.39, Rk. 2]).

Lemma 2.3. (Hoffman, Wielandt). *Let A, B be symmetric $N \times N$ matrices with eigenvalues $\lambda_1^A \le \lambda_2^A \le \ldots \le \lambda_N^A$ and $\lambda_1^B \le \lambda_2^B \le \ldots \le \lambda_N^B$. Then*

$$\sum_{j=1}^N (\lambda_j^A - \lambda_j^B)^2 \le \mathrm{Tr}[(A-B)^2].$$

2.2. **The Proof of Theorem 1.1.** We now proceed to prove Theorem 1.1, using Theorem 1.6. We first prove the second statement of the theorem: let X_N be the matrix ensemble satisfying conditions (1) and (2) of Theorem 1.1, together with the assumption that the entries of $\sqrt{N}X_N$ are bounded by some uniform constant R, $\|\sqrt{N}[X_N]_{ij}\|_\infty \le R$ for all N and all $1 \le i,j \le N$. (This latter assumption subsumes (1).) Denote the blocks of the partition in assumption (2) as $\Pi_N = \{P_1, \ldots, P_r\}$.

Now, let $t = t_N > 0$ (to be chosen later), and let G_N be a GOE_N (with entries of variance $\frac{1}{N}$) independent from X_N. Set

$$\widetilde{X}_N = X_N + tG_N. \tag{2.1}$$

For $1 \le k \le r$, let $\mathbf{Y}_k$ denote the random vector in $\mathbb{R}^{|P_k|}$ given by the entries $[X_N]_{ij}$ with $(i,j) \in P_k$; similarly, let $\mathbf{Z}_k$ be the corresponding entries of G_N. Notice that $\sqrt{N}\mathbf{Y}_k$ is a bounded random vector: by assumption, all of its entries have L^∞-norm $\le R$, and so $\||N|\mathbf{Y}_k|\|_\infty^2 \le R|P_k|^{1/2} \le Rd_N^{1/2}$. The vector $\sqrt{N}\mathbf{Z}_k$ is a standard normal random vector in $\mathbb{R}^{|P_k|}$. Thus, by Theorem 1.6, the law of $\sqrt{N}(\mathbf{Y}_k + t\mathbf{Z}_k)$ satisfies a log-Sobolev inequality with constant

$$c(t) \le \left(K_1 d_N + K_2 \frac{R^2 d_N}{t}\right) R^2 d_N \exp\left(\frac{4R^2 d_N}{t}\right) \le \frac{K_3 R^4 d_N^2}{t} \exp\left(\frac{4R^2 d_N}{t}\right) \tag{2.2}$$

where $K_3 = \max\{K_1, K_2\}$, and we have assumed that $t \le 1$ and $R \ge 1$. By assumption, the random variables $\{\mathbf{Y}_k\}_{k=1}^r$ are independent, as are $\{\mathbf{Z}_k\}_{k=1}^r$. Hence $\{\sqrt{N}(\mathbf{Y}_k + t\mathbf{Z}_k)\}_{k=1}^r$ are independent. Thus, the joint law of entries of $\sqrt{N}\widetilde{X}_N$ is the product measure of the laws of these random variables. As all their laws satisfy log-Sobolev inequalities with the same constant $c(t)$ in (2.2), Segal's Lemma 2.2 shows that:

Corollary 2.4. *The joint law of entries of $\sqrt{N}\widetilde{X}_N$ satisfies a log-Sobolev inequality with constant $c(t)$ of (2.2).*

In particular, Guionnet's Theorem 2.1 shows that the (Lipschitz) linear statistics of the ensemble $\widetilde{X}_N$ are highly concentrated around their means (for fixed t).

Our goal is now to compare the linear statistics of X_N to those of $\widetilde{X}_N$. As usual, let μ_N denote the ESD of X_N, and let $\widetilde{\mu}_N$ denote the ESD of $\widetilde{X}_N$. Then, for each $\epsilon > 0$, and each test function f, we have the following standard triangle inequality estimate.

$$\left|\int f\,d\mu_N - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \le \left|\int f\,d\mu_N - \int f\,d\widetilde{\mu}_N\right| \tag{2.3}$$

$$+ \left|\int f\,d\widetilde{\mu}_N - \mathbb{E}\left(\int f\,d\widetilde{\mu}_N\right)\right| \tag{2.4}$$

$$+ \left|\mathbb{E}\left(\int f\,d\widetilde{\mu}_N\right) - \mathbb{E}\left(\int f\,d\mu_N\right)\right|. \tag{2.5}$$

We will now show that, with a judicious choice of $t = t_N$, each of the quantities (2.3)-(2.5) converges to 0 a.s. We do this in the following three lemmas.

Lemma 2.5. *Let $t = t_N > 0$ be a sequence tending to 0. Then for each $f \in \mathrm{Lip}(\mathbb{R})$,*

$$\left|\int f\,d\mu_N - \int f\,d\widetilde{\mu}_N\right| \to 0 \text{ a.s. as } N \to \infty.$$

Proof. Let $\lambda_1^N \le \lambda_2^N \le \ldots \le \lambda_N^N$ and $\widetilde{\lambda}_1^N \le \widetilde{\lambda}_2^N \le \ldots \le \widetilde{\lambda}_N^N$ be the eigenvalues of X_N and $\widetilde{X}_N$. Then by the Cauchy-Schwarz inequality and Lemma 2.3,

$$\begin{aligned}\left|\int f\,d\mu_N - \int f\,d\widetilde{\mu}_N\right| = \frac{1}{N}\left|\sum_{j=1}^N [f(\lambda_j^N) - f(\widetilde{\lambda}_j^N)]\right| &\le \frac{1}{N}\sum_{i=1}^N \|f\|_{\mathrm{Lip}}\left|\lambda_i^N - \widetilde{\lambda}_i^N\right| \\ &\le \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\sum_{i=1}^N (\lambda_i^N - \widetilde{\lambda}_i^N)^2\right)^{1/2} \\ &\le \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2]\right)^{1/2}.\end{aligned}$$

Now, for any symmetric $N \times N$ matrix A, $\left(\frac{1}{N}\mathrm{Tr}(A^2)\right)^{1/2}$ is the non-commutative L^2-norm of A with respect to the faithful normal state $\frac{1}{N}\mathrm{Tr}$; it is bounded above by the operator norm of A. Applying this to $A = X_N - \widetilde{X}_N = t_N G_N$, we therefore have

$$\left|\int f\,d\mu_N - \int f\,d\widetilde{\mu}_N\right| \le \|f\|_{\mathrm{Lip}}\|X_N - \widetilde{X}_N\|_{\mathrm{op}} = t_N\|f\|_{\mathrm{Lip}}\|G_N\|_{\mathrm{op}} \text{ a.s.}$$

According to [4], the largest eigenvalue $\|G_N\|_{\mathrm{op}}$ of the GOE_N is a.s ≤ 3 for all sufficiently large N. This proves the result. □

Lemma 2.6. *Let $f \in \mathrm{Lip}(\mathbb{R})$, and suppose t_N is chosen such that $c(t_N) = o(\frac{N^2}{\log N})$, where $c(t)$ denote the log-Sobolev constant in* (2.2). *Then*

$$\left|\int f\,d\widetilde{\mu}_N - \mathbb{E}\left(\int f\,d\widetilde{\mu}_N\right)\right| \to 0 \textit{ a.s. as } N \to \infty.$$

Proof. Theorem 2.1 and Corollary 2.4 yield that, for any $\epsilon > 0$ ans $N \in \mathbb{N}$,

$$\mathbb{P}\left(\left|\int f\,d\widetilde{\mu}_N - \mathbb{E}\left(\int f\,d\widetilde{\mu}_N\right)\right| \ge \epsilon\right) \le 2\exp\left(-\frac{N^2\epsilon^2}{c(t_N)\|f\|_{\mathrm{Lip}}^2}\right).$$

By assumption, there is a sequence $s_N \to 0$ so that $c(t_N) = \frac{N^2}{\log N}s_N$. Thus

$$\exp\left(-\frac{N^2\epsilon^2}{c(t_N)\|f\|_{\mathrm{Lip}}^2}\right) = \exp\left(-\frac{\epsilon^2}{\|f\|_{\mathrm{Lip}}^2}\frac{\log N}{s_N}\right) = N^{-\frac{\epsilon^2}{\|f\|_{\mathrm{Lip}}^2}\frac{1}{s_N}}.$$

Since $\frac{1}{s_N} \to \infty$, for all sufficiently large N this is $\le \frac{1}{N^2}$. The result now follows from the Borel–Cantelli lemma. □

Lemma 2.7. *Let $t = t_N > 0$ be a sequence tending to* 0*. Then for each $f \in \mathrm{Lip}(\mathbb{R})$,*

$$\left|\mathbb{E}\left(\int f\,d\widetilde{\mu}_N\right) - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \to 0 \textit{ as } N \to \infty.$$

Proof. In Lemma 2.5, we showed that $\int f\,d\widetilde{\mu}_N - \int f\,d\mu_N \to 0$ a.s. Hence, to show that the expectation goes to 0, it suffices to show that these random variables have finite L^1-norm for all large N. This follows from estimates like the ones in the proof of Lemma 2.5:

$$\begin{aligned}\mathbb{E}\left(\left|\int f\,d\widetilde{\mu}_N - \int f\,d\mu_N\right|\right) &\le \mathbb{E}\left(\frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2]\right)^{1/2}\right) \\ &\le \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\mathbb{E}\left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2]\right)\right)^{1/2} = \|f\|_{\mathrm{Lip}}t_N^{1/2} < \infty\end{aligned}$$

where we applied Jensen's inequality in the second step. The result follows. □

We can now prove the theorem under the boundedness assumption.

Proof of Theorem 1.1 Assuming $\sqrt{N}X_N$ has entries uniformly bounded by R. In light of Lemma 2.5-2.7, it suffices to show that there is a sequence $t_N > 0$ such that $t_N \to 0$ and $c(t_N) = o(\frac{N^2}{\log N})$. For N sufficiently large, we define

$$t_N := \frac{5R^2 d_N}{\log \frac{N}{K_3 R^2}}.$$

By Assumption (2) of Theorem 1.1, $d_N = o(\log N)$, and hence $t_N \to 0$ as $N \to \infty$. As such, $\frac{R^2}{t_N} > 2$ for all large N, and it follows (from elementary calculus) that $\frac{R^2}{t_N} d_N^2 \le \exp(\frac{R^2}{t_N} d_N)$. Thus, (2.2) yields

$$c(t_N) \le K_3 R^2 \cdot \frac{R^2}{t_N} d_N^2 \exp\left(\frac{4R^2 d_N}{t_N}\right) \le K_3 R^2 \exp\left(\frac{5R^2 d_N}{t_N}\right) = N = o\left(\frac{N^2}{\log N}\right).$$

This concludes the proof. □

Remark 2.8. We could have arranged for $c(t_N)$ to be of larger order but still $o(N^2/\log N)$, but this would only have resulted in the ratio d_N/t_N being a constant factor larger, and thus would still require $d_N = o(\log N)$ in order for it to be possible for $t_N \to 0$. Moreover, even if the blunt estimate $\frac{R^2}{t_N} d_N^2 \le \exp(\frac{R^2}{t_N} d_N)$ had not been employed, or even if (2.2) were known to hold without the prefactor (as might be true if the sharp form Theorem 1.6 held with a constant independent of dimension), it would still be impossible to arrange for $t_N \to 0$ while $c(t_N) = o(\frac{N^2}{\log N})$ unless $d_N = o(\log N)$. That is: the result of Theorem 1.1 cannot be improved using the approach of this paper.

To conclude the proof, it remains only to remove the boundedness assumption on the entries of $\sqrt{N}X_N$ (at the expense of a downgrade from almost sure convergence to convergence in probability). This is where the uniform integrability comes in, via a standard cutoff argument that we briefly outline. Let $\epsilon, \eta > 0$. Let $f \in \mathrm{Lip}(\mathbb{R})$. By uniform integrability, there exists some $R \ge 0$ such that

$$\mathbb{E}\left(N[X_N]_{ij}^2 \cdot \mathbb{1}_{\{\sqrt{N}|[X_N]_{ij}|>R\}}\right) < \min(1,\eta) \cdot \epsilon^2/(9\|f\|_{\mathrm{Lip}}^2)$$

for all i, j, N. Let $\widehat{X}_N$ be the matrix whose entries are the appropriate cutoffs of X_N:

$$[\widehat{X}_N]_{ij} = [X_N]_{ij} \cdot \mathbb{1}_{\{\sqrt{N}|[X_N]_{ij}|\le R\}}.$$

Then $\|\sqrt{N}\widehat{X}_{ij}\|_\infty \le R$ for all N, i, j. Let $\widehat{\mu}_N$ denote the ESD of $\widehat{X}_N$. The preceding proof shows that $\int f\, d\widehat{\mu}_N$ converge to its mean almost surely, and hence in probability. We now compare the linear statistics of μ_N and $\widehat{\mu}_N$. This is similar to the preceding analysis. We make the standard $\epsilon/3$-decomposition:

$$\begin{aligned}\mathbb{P}\left(\left|\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right)\right| \ge \epsilon\right) \le\ & \mathbb{P}\left(\left|\int f\, d\mu_N - \int f\, d\widehat{\mu}_N\right| \ge \frac{\epsilon}{3}\right) \\ &+ \mathbb{P}\left(\left|\int f\, d\widehat{\mu}_N - \mathbb{E}\left(\int f\, d\widehat{\mu}_N\right)\right| \ge \frac{\epsilon}{3}\right) \\ &+ \mathbb{P}\left(\left|\mathbb{E}\left(\int f\, d\widehat{\mu}_N\right) - \mathbb{E}\left(\int f\, d\mu_N\right)\right| \ge \frac{\epsilon}{3}\right).\end{aligned} \tag{2.6}$$

The above proof in the uniform bounded case shows that the second term in (2.6) converges to 0 as $N \to \infty$. The first term on the right hand side of (2.6) is bounded using the same reasoning as done in the proof of Lemma

2.5:

$$\mathbb{P}\left(\left|\int f\,d\mu_N - \int f\,d\widehat{\mu}_N\right| \geq \frac{\epsilon}{3}\right) \leq \frac{9\|f\|_{\mathrm{Lip}}^2}{\epsilon^2 N} \sum_{1\leq i,j\leq N} \mathbb{E}\left(([X_N]_{ij} - [\widehat{X}_N]_{ij})^2\right)$$
$$= \frac{9\|f\|_{\mathrm{Lip}}^2}{\epsilon^2 N} \sum_{1\leq i,j\leq N} \mathbb{E}\left([X_N]_{ij}^2 \cdot \mathbb{1}_{\{\sqrt{N}|[X_N]_{ij}|>R\}}\right) < \eta.$$

Finally, the third term is bounded as in Lemma 2.7:

$$\left|\mathbb{E}\left(\int f\,d\widehat{\mu}_N\right) - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\mathbb{E}\left(\mathrm{Tr}[(X_N - \widehat{X}_N)^2]\right)\right)^{1/2}$$
$$= \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\sum_{1\leq i,j\leq N} \mathbb{E}\left([X_n]_{ij}^2 \cdot \mathbb{1}_{\{\sqrt{N}|[X_N]_{ij}|>R\}}\right)\right)^{1/2} < \frac{\epsilon}{3},$$

so $\mathbb{P}\left(\left|\mathbb{E}\left(\int f\,d\widehat{\mu}_N\right) - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \geq \frac{\epsilon}{3}\right) = 0$. Therefore

$$\limsup_{N\to\infty} \mathbb{P}\left(\left|\int f\,d\mu_N - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \geq \epsilon\right) \leq \eta.$$

Since $\eta > 0$ was arbitrary, we have $\mathbb{P}\left(\left|\int f\,d\mu_N - \mathbb{E}\left(\int f\,d\mu_N\right)\right| \geq \epsilon\right) \to 0$ as $N \to \infty$, giving convergence in probability. This concludes the proof.

Remark 2.9. Instead of doing the Gaussian mollification and *then* the cutoff argument, we could combine the two in the hopes of proving almost sure convergence in the general case. The obstruction to this is Lemma 2.5, where we used the fact (proved in [4]) that the GOE_N has no asymptotic outlier eigenvalues above 2: with probability 1, all eigenvalues are eventually ≤ 3, for example. If we were to combine the cutoff argument with the mollification argument, in this lemma $X_N - \widetilde{X}_N$ would not be $t_N G_N$ but rather $t_N G_N + (X_N - \widehat{X}_N)$; i.e. Gaussian noise plus a matrix whose entries are of the form $[X_N]_{ij}\mathbb{1}_{\sqrt{N}|[X_N]_{ij}|>R}$. If the entries of X_N were independent, then additional moment growth assumptions would imply the necessary lack of outlier eigenvalues following [4]; however, in our case where the entries may be correlated, the behavior of the largest eigenvalue is, at present, unknown.

2.3. **Theorem 1.2, a Generalization, and Applications.** We begin with a lemma which appeared in the second author's paper [44, Prop. 6]. We reproduce the simple proof here, for completeness.

Lemma 2.10. *Let* $\mathbf{X}$ *be a random vector in* $\mathbb{R}^d$ *whose law satisfies a log-Sobolev inequality* (1.4) *with constant* c*. Let* $T\colon \mathbb{R}^d \to \mathbb{R}^d$ *be a Lipschitz map. Then the law of* $T(\mathbf{X})$ *satisfies a log-Sobolev inequality with constant* $c\|T\|_{\mathrm{Lip}}^2$.

Proof. Let μ denote the law of $\mathbf{X}$. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a locally-Lipschitz non-negative function. Then $f\circ T$ is locally-Lipschitz and non-negative. Since μ satisfies the LSI with constant c, it follows that

$$\int (f\circ T)^2 \log\frac{(f\circ T)^2}{\int (f\circ T)^2}\,d\mu \leq c\int |\nabla(f\circ T)|^2\,d\mu. \tag{2.7}$$

Since T is Lipschitz, we also have the pointwise estimate

$$|\nabla(f\circ T)| \leq (|\nabla f|\circ T)\|T\|_{\mathrm{Lip}}.$$

By a change of variables, (2.7) therefore shows that

$$\int f^2 \log\frac{f^2}{\int f^2\,dT_*\mu}\,dT_*\mu \leq c\|T\|_{\mathrm{Lip}}^2 \int |\nabla f|^2\,dT_*\mu.$$

Thus, the push-forward measure $T_*\mu$ satisfies the LSI with constant $c\|T\|_{\mathrm{Lip}}^2$. Since $T_*\mu$ is the law of $T(\mathbf{X})$, this concludes the proof. ∎

The following theorem covers a wide range of examples of correlated random matrix ensembles. We use the notation $\mathbb{M}_N^{\mathrm{sym}}$ to denote the vector space of real $N \times N$ symmetric matrices, equipped with the Hilbert–Schmidt inner product.

Theorem 2.11. *Let $\{\xi_{ij} \colon 1 \le i \le j\}$ be a triangular array of i.i.d. random variables whose common law satisfies a log-Sobolev inequality* (1.4)*. Let Ξ_N be the $N \times N$ symmetric random matrix with entries $[\Xi_N]_{ij} = \xi_{ij}$ for $1 \le i \le j \le N$. Let $T_N \colon \mathbb{M}_N^{\mathrm{sym}} \to \mathbb{M}_N^{\mathrm{sym}}$ be a Lipschitz function, with $\|T_N\|_{\mathrm{Lip}} = o(\frac{N}{\sqrt{\log N}})$.*

Let $X_N = T_N(N^{-1/2}\Xi_N)$, and let μ_N denote the ESD of X_N. Then μ_N converges to its mean almost surely:

$$\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right) \to 0 \ a.s. \qquad \text{for all } \ f \in \mathrm{Lip}(\mathbb{R}). \tag{2.8}$$

Proof. Let $S_N \colon \mathbb{M}_N^{\mathrm{sym}} \to \mathbb{M}_N^{\mathrm{sym}}$ be the conjugate-scaled map $S_N(A) = N^{1/2}T_N(N^{-1/2}A)$. Then S_N is also Lipschitz with $\|S_N\|_{\mathrm{Lip}} = \|T_N\|_{\mathrm{Lip}}$. By assumption, the entries ξ_{ij} satisfy a LSI with some constant c; by Lemma 2.2, the joint law of Ξ_N therefore satisfies the LSI with constant c. Hence, by Lemma 2.10, $S_N(\Xi_N) = \sqrt{N}X_N$ satisfies the LSI with constant $c\|S_N\|_{\mathrm{Lip}}^2 = c\|T_N\|_{\mathrm{Lip}}^2$. By Theorem 2.1, it therefore follows that, for any $\epsilon > 0$ and $f \in \mathrm{Lip}(\mathbb{R})$,

$$\mathbb{P}\left(\left|\int f\, d\mu_N - \mathbb{E}\left(\int f\, d\mu_N\right)\right| \ge \epsilon\right) \le 2\exp\left(-\frac{N^2\epsilon^2}{c\|T_N\|_{\mathrm{Lip}}^2 ||f||_{\mathrm{Lip}}^2}\right).$$

By assumption, $\|T_N\|_{\mathrm{Lip}}^2 = o(\frac{N^2}{\log N})$. The result now follows exactly as in the proof of Lemma 2.6. $\square$

We now prove Theorem 1.2, essentially as a Corollary to Theorem 2.11 (although we really prove it as a corollary to *the proof* of Theorem 2.11, to most easily deduce the optimal result).

Proof of Theorem 1.2. To begin, we clarify what is meant by "jointly Gaussian". We say a random vector $\mathbf{X} \in \mathbb{R}^d$ has jointly Gaussian entries if there is an affine map $T \colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\mathbf{X} = T(\mathbf{G})$, where $\mathbf{G}$ has i.i.d. normal entries. A more standard definition of "jointly Gaussian" — that the joint law of the centered entries should have a density of the form $c\exp(\mathbf{x} \cdot C^{-1}\mathbf{x})$ for a positive definite matrix C and a normalization constant c — is a special case: it is easy to check $\sqrt{C^{-1}}(\mathbf{X})$ has i.i.d. standard normal entries, and so $T = \sqrt{C}$ will suffice.

Let $\Pi_N = \{P_1, \ldots, P_r\}$ denote the partition of $\{(i,j) \colon 1 \le i \le j \le N\}$ in the theorem, and for $1 \le k \le r$ let $\mathbf{X}_k$ denote the random vector given by the entries of X_N with indices in P_k. By assumption, the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_r$ are independent; it follows that there are affine maps $T_1, \ldots, T_r$ with $T_k \colon \mathbb{R}^{|P_k|} \to \mathbb{R}^{|P_k|}$, such that $N^{1/2}\mathbf{X}_k = T_k(\mathbf{G}_k)$, where $\mathbf{G}_k$ is a standard Gaussian random vector in $\mathbb{R}^{|P_k|}$. The law of $\mathbf{G}_k$ satisfies a log-Sobolev inequality with constant 1 (cf. [24]), and therefore by Lemma 2.10, the law of $N^{1/2}\mathbf{X}_k$ satisfies a log-Sobolev inequality with constant $\|T_k\|_{\mathrm{Lip}}^2$.

Now, T_k has the form $T_k = \mathring{T}_k + N^{1/2}\mathbb{E}(\mathbf{X}_k)$ for some linear map $\mathring{T}_k$, and $\|T_k\|_{\mathrm{Lip}} = \|\mathring{T}_k\|_{\mathrm{op}} \le \|\mathring{T}_k\|_{\mathrm{HS}}$, where $\|\cdot\|_{\mathrm{HS}}$ is the (un-normalized) Hilbert–Schmidt norm. Thus, we have

$$\|T_k\|_{\mathrm{Lip}}^2 \le \sum_{a,b=1}^{|P_k|} [\mathring{T}_k]_{ab}^2 \tag{2.9}$$

where we use the indices a, b to enumerate the entries of $\mathbf{X}_k$. Now, note that

$$\mathrm{Var}(N^{1/2}[\mathbf{X}_k]_a) = \mathrm{Var}([\mathring{T}_k(\mathbf{G}_k)]_a) = \mathrm{Var}\left(\sum_b [\mathring{T}_k]_{ab}[\mathbf{G}_k]_b\right) = \sum_b [\mathring{T}_k]_{ab}^2$$

because $\mathrm{Cov}([\mathbf{G}_k]_b, [\mathbf{G}_k]_c) = \delta_{bc}$. By assumption, there is a uniform bound R so that $\mathrm{Var}(N^{1/2}[\mathbf{X}_k]_a) \le N\mathbb{E}([\mathbf{X}_k]_a^2) \le R^2$ for all k and a. Thus (2.9) yields

$$\|T_k\|_{\mathrm{Lip}}^2 \le \sum_{a=1}^{|P_k|} \mathrm{Var}(N^{1/2}[\mathbf{X}_k]_a) \le R^2|P_k| \le R^2 d_N.$$

We have thus shown that the law of $N^{1/2}\mathbf{X}_k$ satisfies a log-Sobolev inequality with constant $R^2 d_N$, for each k. Since the random variables $\mathbf{X}_k$ are independent, Lemma 2.2 shows that the joint law of entries of $N^{1/2}X_N$ satisfies a log-Sobolev inequality with constant $R^2 d_N$. Since $d_N = o(\frac{N^2}{\log N})$, the almost sure convergence now follows precisely as in the proof of Theorem 2.11 above. $\square$

Remark 2.12. (1) Theorem 1.2 is really a special case of Theorem 2.11; what the preceding proof essentially does is show that the affine function which is "block diagonal" combining all the block T_k maps has Lipschitz norm $= o(\frac{N}{\sqrt{\log N}})$.

(2) The proof shows that it is really enough to assume, in the statement of Theorem 1.2, that the scaled *variances* of the entries are uniformly bounded. However, in any instance we wish to apply the theorem, we must have the expectations of empirical integrals converging, and hence there is no loss in making the nominally stronger assumption that the scaled second *moments* of the entries are uniformly bounded.

2.4. **Theorem 1.4 and Generalizations.** In this section, we begin by showing how to prove Theorem 1.4 as a straightforward corollary to Theorems 1.1 and 2.11. To begin, we note that the topic of the paper [34] is the convergence in expectation of ensembles of this form (and slightly more general forms). In particular, using the tools of operator-valued free probability, Shlyakhtenko showed that all ensembles of this form have a limiting ESD, that can be computed (in principle) in terms of the spectral measure of the operator η on $L^\infty[0,1]$ defined by $\eta(f)(x) = \int_0^1 f(y)g(x,y)^2\,dy$ (embedded into a Fock space type model). The limiting ESD can be computed exactly in many cases; in particular, if $\int_0^1 g(x,y)\,dy = 1$ for each x, then the limit law is semicircular; cf. [34, Remark 3.8]. As such, we concern ourselves here only with the question of upgrading from convergence in expectation to convergence in probability / almost sure convergence, where appropriate.

Proof of Theorem 1.4. We apply Theorem 1.1 to the ensemble X_N. The upper-triangular entries of X_N are all independent, and so condition (2) of Theorem 1.1 (on the size of independent blocks) is automatically satisfied. Hence, to conclude convergence of the ESD to its mean in probability, it suffices to show that the family $\{N[X_N]_{ij}^2\}_{N\in\mathbb{N},1\le i,j\le N}$ is uniformly integrable. Note that

$$N[X_N]_{ij}^2 = g(i/N,j/N)\xi_{ij}^2.$$

By assumption $g \in C([0,1]^2)$, and so g is bounded. Thus $\mathbb{E}(N[X_N]_{ij}^2) \le \|g\|_\infty \mathbb{E}(\xi_{ij}^2) = \|g\|_\infty$ for all i,j,N (since the ξ_{ij} are presumed to be centered with variance 1). Thus, there is a uniform bound on the expectation of each element in this family of nonnegative random variables, and it follows that the family is uniformly integrable. Thus, by Theorem 1.1, the ESD of X_N converges to its mean in probability.

For the second statement of the theorem: first, if ξ_{ij} are bounded random variables $\le R$, then the entries $\sqrt{N}|[X_N]_{ij}| \le \|g\|_\infty R$ are uniformly bounded, and so almost sure convergence follows from the last statement of Theorem 1.1. In the case where we assume the law of the ξ_{ij} satisfies a log-Sobolev inequality, we apply Theorem 2.11. Note that our ensemble has the form $T_N(N^{-1/2}\Xi_N) = N^{-1/2}T_N(\Xi_N)$, where Ξ_N has entries ξ_{ij} and T_N is the linear "diagonal" map

$$[T_N(\Xi_N)]_{ij} = g(i/N,j/N)[\Xi_N]_{ij}.$$

Since the entries ξ_{ij} are i.i.d. and are assumed to satisfy a log-Sobolev inequality, to establish almost sure convergence of the ESD, it suffices by Theorem 2.11 to show that $\|T_N\|_{\mathrm{Lip}} = o(\frac{N}{\sqrt{\log N}})$. But since T_N is linear and diagonal, its Lipschitz norm (i.e. operator norm) is simply the maximum modulus of the entries, $\|T_N\|_{\mathrm{op}} = \max_{i,j}|g(i/N,j/N)| \le \|g\|_\infty = O(1) = o(\frac{N}{\sqrt{\log N}})$. This concludes the proof. $\square$

3. Mollified Log-Sobolev Inequalities on $\mathbb{R}^d$

In this section we will prove Theorem 1.6. For convenience, we restate it below as Theorem 3.1, in measure theoretic language.

Theorem 3.1. *Let μ be a probability measure on $\mathbb{R}^d$ whose support is contained in a ball of radius R, and let γ_t be the centered Gaussian of variance t with $0 < t \le R^2$, i.e., $\gamma_t(x) = (2\pi t)^{-d/2} \exp(-\frac{|x|^2}{2t})\, dx$. Then for some absolute constant K, the optimal log-Sobolev constant $c(t)$ for the convolution $\mu * \gamma_t$ satisfies*

$$c(t) \le K\, R^2 \exp\left(20d + \frac{5R^2}{t}\right).$$

K can be taken above to be 289.

Remark 3.2. Theorem 3.1 is slightly more general than Theorem 1.6, since it only requires the support to be contained in some ball of radius R; by contrast, in Theorem 1.6, R is the radius of a ball *centered at* 0 containing $\operatorname{supp}\mu$. If we use the theorem in this form, we could actually improve Theorem 1.1 by softening the requirement that the entires be uniformly square integrable, only requiring their centered versions $\sqrt{N}([X_N]_{ij} - \mathbb{E}([X_N]_{ij}))$ to be uniformly square integrable. However, since any ensembles we wish to apply Theorem 1.1 to must converge in expectation, this does not given any practical improvement.

3.1. **The Proof of Theorem 3.1.** To prove Theorem 3.1, we use the following theorem (see [19, p.288, Thm. 1.2]):

Theorem 3.3. (Cattiaux, Guillin, Wu). *Let μ be a probability measure on $\mathbb{R}^d$ with $d\mu(x) = e^{-V(x)}dx$ for some $V \in C^2(\mathbb{R}^d)$. Suppose the following:*

(1) There exists a constant $K \le 0$ such that $\operatorname{Hess}(V) \ge KI$.
(2) There exists a $W \in C^2(\mathbb{R}^d)$ with $W \ge 1$ and constants $b, c > 0$ such that

$$tW(x) - \langle \nabla V, \nabla W\rangle(x) \le (b - c|x|^2)W(x)$$

for all $x \in \mathbb{R}^d$.

Then μ satisfies a LSI.

In particular, let $r_0, b', \lambda > 0$ be such that

$$tW(x) - \langle \nabla V, \nabla W\rangle(x) \le -\lambda W(x) + b' \mathbb{1}_{B_{r_0}}$$

where B_{r_0} denotes the ball centered at 0 of radius r_0 (the existence of such r_0, b', λ is implied by Assumption 2). By [7, p.61, Thm. 1.4]*, μ satisfies a Poincaré inequality with constant C_P; that is, for every sufficiently smooth g with $\int g\, d\mu = 0$,*

$$\int g^2 d\mu \le C_P \int |\nabla g|^2 d\mu; \tag{3.1}$$

C_P can be taken to be $(1 + b'\kappa_{r_0})/\lambda$, where κ_{r_0} is the Poincaré constant of μ restricted to B_{r_0}. A bound for κ_{r_0} is

$$\kappa_{r_0} \le D r_0^2 \frac{\sup_{x\in B_{r_0}} p(x)}{\inf_{x \in B_{r_0}} p(x)},$$

where $p(x) = e^{-V(x)}$ and D is some absolute constant that can be taken to be $4/\pi^2$. Let

$$\begin{aligned} A =& \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right) + \epsilon \\ B =& \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right)\left(b + c\int |x|^2 d\mu(x)\right), \end{aligned}$$

where ϵ is an arbitrarily chosen parameter. Then μ satisfies a LSI with constant $A + (B+2)C_P$.

We remark that the statement of Theorem 3.3 is given in [19] in the more general context of Riemannian manifolds. Also, the constants given above are derived in [19] but not presented there; for our purposes we have collected those constants and presented them here.

With the above, we now prove Theorem 3.1, which we restate here for the reader's convenience.

Theorem 3.4. *Let μ be a probability measure on $\mathbb{R}^d$ whose support is contained in a ball of radius R, and let γ_t be the centered Gaussian of variance t with $0 < t \leq R^2$, i.e., $d\gamma_t(x) = (2\pi t)^{-n/2} \exp(-\frac{|x|^2}{2t})dx$. Then for some absolute constant K, the optimal log-Sobolev constant $c(t)$ for $\mu * \gamma_t$ satisfies*

$$c(t) \leq K\, R^2 \exp\left(20n + \frac{5R^2}{t}\right).$$

K can be taken above to be 289.

Proof. By translation invariance of LSI, we will assume that μ is supported in B_R. We will apply Theorem 3.3 to μ_t and compute the appropriate bounds and expressions for K, W, b, c, r_0, b', λ, κ_{r_0}, C_P, $\int |x|^2 d\mu_t(x)$, A, and B.

To find K, b, and c, we follow the computations as done in [39, pp. 7-8]. Let $V(x) = \frac{x^2}{2t}$ and $V_t(x) = -\log(p_t(x))$, so

$$d\mu_t(x) = e^{-V_t(x)}dx = d(e^{-V} * \mu)(x).$$

Also let

$$d\mu_x(z) = \frac{1}{p_t(x)} e^{-V(x-z)} d\mu(z),$$

so μ_x is a probability measure for each $x \in \mathbb{R}^d$. Then for $X \in \mathbb{R}^d$ with $|X| = 1$,

$$\begin{aligned}
\mathrm{Hess}(V_t)(X,X)(x) =& \left(\int_{B_R} \nabla_X V(x-z) d\mu_x(z)\right)^2 - \int_{B_R} \left(|\nabla_X V(x-z)|^2 - \mathrm{Hess}(V)(X,X)(x-z)\right) d\mu_x(z) \\
=& \frac{1}{t} - \left(\int_{B_R} |\nabla_X V(x-z)|^2 d\mu_x(z) - \left(\int_{B_R} \nabla_X V(x-z) d\mu_x(z)\right)^2\right) \\
& \text{since } \mathrm{Hess}(V) = \frac{1}{t} I.
\end{aligned}$$

But for any C^1 function f,

$$\begin{aligned}
\int_{B_R} f^2 d\mu_x(z) - \left(\int_{B_R} f\, d\mu_x(z)\right)^2 =& \frac{1}{2} \int_{B_R \times B_R} (f(z) - f(y))^2 d\mu_x(z) d\mu_x(y) \\
\leq& 2R^2 \sup |\nabla f|^2,
\end{aligned}$$

so for $f = \nabla_X V$, we get

$$\mathrm{Hess}(V_t)(X,X)(x) \geq \frac{1}{t} - 2R^2 \sup |\nabla(\nabla_X V)|^2 = \frac{1}{t} - \frac{2R^2}{t^2}.$$

So we take

$$K = \frac{1}{t} - \frac{2R^2}{t^2}.$$

Note $K \leq 0$ since $t \leq R^2$.

Let

$$W(x) = \exp\left(\frac{|x|^2}{16t}\right).$$

Then

$$\begin{aligned}\frac{tW-\langle\nabla V_t,\nabla W\rangle}{W}(x)=&\frac{n}{8t}+\frac{|x|^2}{64t^2}-\frac{1}{16t}\int_{B_R}\langle x,\nabla V(x-z)\rangle d\mu_x(z)\\=&\frac{n}{8t}+\frac{|x|^2}{64t^2}-\frac{1}{16t^2}\int_{B_R}\left(|x|^2-\langle x,z\rangle\right)d\mu_x(z)\\\leq&\frac{n}{8t}-\frac{3|x|^2}{64t^2}+\frac{1}{16t^2}\sup_{z\in B_R}\langle x,z\rangle\\=&\frac{n}{8t}-\frac{3|x|^2}{64t^2}+\frac{1}{16t^2}R|x|.\end{aligned}$$

Using $|x|\leq|x|^2/2R+R/2$ above, we get

$$\frac{tW-\langle\nabla V_t,\nabla W\rangle}{W}(x)\leq\frac{n}{8t}-\frac{3|x|^2}{64t^2}+\frac{1}{16t^2}R\left(\frac{|x|^2}{2R}+\frac{R}{2}\right)=\frac{n}{8t}+\frac{R^2}{32t^2}-\frac{1}{64t^2}|x|^2,$$

so we take

$$\begin{aligned}b=&\frac{n}{8t}+\frac{R^2}{32t^2},\\c=&\frac{1}{64t^2}.\end{aligned}$$

Now let

$$\begin{aligned}r_0=&\sqrt{16nt+2R^2},\\b'=&\frac{1}{4t}\exp\left(n+\frac{R^2}{8t}-1\right),\\\lambda=&\frac{n}{8t}.\end{aligned}$$

We claim that

$$b-c|x|^2\leq-\lambda+b'\exp\left(-\frac{|x|^2}{16t}\right)\mathbb{1}_{B_{r_0}},\qquad\text{i.e.,}\qquad\frac{b+\lambda-c|x|^2}{b'}\exp\left(\frac{|x|^2}{16t}\right)\leq\mathbb{1}_{B_{r_0}},$$

so that

$$tW(x)-\langle\nabla V,\nabla W\rangle(x)\leq-\lambda W(x)+b'\mathbb{1}_{B_{r_0}}.$$

We have

$$\begin{aligned}\frac{b+\lambda-c|x|^2}{b'}\exp\left(\frac{|x|^2}{16t}\right)=&4t\exp\left(-n-\frac{R^2}{8t}+1\right)\left(\frac{n}{8t}+\frac{R^2}{32t^2}+\frac{n}{8t}-\frac{|x|^2}{64t^2}\right)\exp\left(\frac{|x|^2}{16t}\right)\\=&\left(n+\frac{R^2}{8t}-\frac{|x|^2}{16t}\right)\exp\left(-\left(n+\frac{R^2}{8t}-\frac{|x|^2}{16t}\right)+1\right).\end{aligned}$$

For $|x|\geq r_0$, the above expression is nonpositive, and for $|x|\leq r_0$, the above expression is of the form ue^{-u+1}, which has a maximum value of 1, as desired.

Now we estimate κ_{r_0} by estimating $\sup_{x\in B_{r_0}}p_t(x)$ and $\inf_{x\in B_{r_0}}p_t(x)$. For $x\in B_{r_0}$, we have

$$p_t(x)=\int_{B_R}(2\pi t)^{-n/2}\exp\left(-\frac{|x-y|^2}{2t}\right)d\mu(y)\leq\int_{B_R}(2\pi t)^{-n/2}d\mu(y)=(2\pi t)^{-n/2}$$

and

$$p_t(x) = \int_{B_R} (2\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2t}\right) d\mu(y) \geq \int_{B_R} (2\pi t)^{-n/2} \exp\left(-\frac{(r_0+R)^2}{2t}\right) d\mu(y)$$
$$= (2\pi t)^{-n/2} \exp\left(-\frac{(r_0+R)^2}{2t}\right),$$

so

$$\kappa_{r_0} \leq D r_0^2 \frac{\sup_{x\in B_{r_0}} p(x)}{\inf_{x\in B_{r_0}} p(x)} \leq D r_0^2 \exp\left(\frac{(r_0+R)^2}{2t}\right).$$

We then take

$$\begin{aligned} C_P =& \frac{1+b'\kappa_{r_0}}{\lambda} \\ \leq& \frac{8t}{n}\left(1+\frac{1}{4t}\exp\left(n+\frac{R^2}{8t}-1\right)\cdot D r_0^2 \exp\left(\frac{(r_0+R)^2}{2t}\right)\right) \\ =& \frac{8t}{n} + \frac{D}{e}\left(32t+\frac{4R^2}{n}\right)\exp\left(n+\frac{R^2}{8t}+\frac{(\sqrt{16nt+2R^2}+R)^2}{2t}\right). \end{aligned}$$

Using $\sqrt{a}+\sqrt{b} \leq \sqrt{2(a+b)}$ and the assumptions $t \leq R^2$ and $n \geq 1$ above, we get

$$\begin{aligned} C_P \leq& \frac{8R^2}{1} + \frac{D}{e}\left(32R^2+\frac{4R^2}{1}\right)\exp\left(n+\frac{R^2}{8t}+\frac{\sqrt{2(16nt+2R^2+R^2)}^2}{2t}\right) \\ =& 8R^2 + \frac{36D}{e}R^2\exp\left(17n+\frac{25R^2}{8t}\right) \\ \leq& \left(8+\frac{36D}{e}\right)R^2\exp\left(17n+\frac{25R^2}{8t}\right). \end{aligned}$$

Next, we estimate $\int |x|^2 d\mu_t(x)$:

$$\begin{aligned} \int_{\mathbb{R}^d} |x|^2 d\mu_t(x) =& \int_{\mathbb{R}^d}\int_{B_R} |x|^2 (2\pi t)^{-n/2}\exp\left(-\frac{|x-y|^2}{2t}\right) d\mu(y)dx \\ =& (2\pi t)^{-n/2}\int_{B_R}\int_{\mathbb{R}^d} |x+y|^2 \exp\left(-\frac{|x|^2}{2t}\right) dx\, d\mu(y) \\ & \text{by replacing } x \to x+y \\ =& (2\pi t)^{-n/2}\int_{B_R}\int_{\mathbb{R}^d} (|x|^2+|y|^2)\exp\left(-\frac{|x|^2}{2t}\right) dx\, d\mu(y) \\ &+ (2\pi t)^{-n/2}\int_{B_R}\int_{\mathbb{R}^d} 2\langle x,y\rangle \exp\left(-\frac{|x|^2}{2t}\right) dx\, d\mu(y). \end{aligned}$$

The second integral in the last expression above equals 0 since the integrand is an odd function of x. So

$$\begin{aligned} \int_{\mathbb{R}^d} |x|^2 d\mu_t(x) =& (2\pi t)^{-n/2}\int_{B_R}\int_{\mathbb{R}^d} (|x|^2+|y|^2)\exp\left(-\frac{|x|^2}{2t}\right) dx\, d\mu(y) \\ \leq& (2\pi t)^{-n/2}\int_{\mathbb{R}^d}\int_{B_R} (|x|^2+R^2)\exp\left(-\frac{|x|^2}{2t}\right) d\mu(y)dx \\ =& (2\pi t)^{-n/2}\int_{\mathbb{R}^d} (|x|^2+R^2)\exp\left(-\frac{|x|^2}{2t}\right) dx \\ =& nt+R^2, \end{aligned}$$

the last integral computed using polar coordinates.

To get expressions for A, B, we choose $\epsilon = 16t$; then A, B satisfy

$$A = \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right) + \epsilon = 128t^2\left(\frac{1}{16t} - \left(\frac{1}{2t} - \frac{R^2}{t^2}\right)\right) + 16t = 128R^2 - 40t \leq 128R^2$$

and

$$\begin{aligned} B = \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right)\left(b + c\int |x|^2 d\mu_t(x)\right) \leq & 128t^2\left(\frac{1}{16t} - \left(\frac{1}{2t} - \frac{R^2}{t^2}\right)\right)\left(\frac{n}{8t} + \frac{R^2}{32t^2} + \frac{1}{64t^2}\left(nt + R^2\right)\right) \\ = & \frac{18nR^2}{t} + \frac{6R^4}{t^2} - \frac{63n}{8} - \frac{21R^2}{8} \\ \leq & \frac{18nR^2}{t} + \frac{6R^4}{t^2} - 2. \end{aligned}$$

Putting everything together, we get that the optimal log-Sobolev constant $c(t)$ for μ_t satisfies

$$\begin{aligned} c(t) \leq & A + (B+2)C_P \\ \leq & 128R^2 + \left(\frac{18nR^2}{t} + \frac{6R^4}{t^2} - 2 + 2\right)\left(8 + \frac{36D}{e}\right)R^2 \exp\left(17n + \frac{25R^2}{8t}\right) \\ = & 128R^2 + 12 \cdot \frac{R^2}{2t}\left(3n + \frac{R^2}{t}\right)\left(8 + \frac{36D}{e}\right)R^2 \exp\left(17n + \frac{25R^2}{8t}\right). \end{aligned}$$

Applying $u \leq e^u$ to two of the terms in the expression above, we get

$$\begin{aligned} c(t) \leq & 128R^2 + 12\exp\left(\frac{R^2}{2t}\right)\exp\left(3n + \frac{R^2}{t}\right)\left(8 + \frac{36D}{e}\right)R^2 \exp\left(17n + \frac{25R^2}{8t}\right) \\ = & 128R^2 + \left(96 + \frac{432D}{e}\right)R^2 \exp\left(20n + \frac{37R^2}{8t}\right) \\ \leq & \left(128 + 96 + \frac{432D}{e}\right)R^2 \exp\left(20n + \frac{5R^2}{t}\right) \\ \leq & 289R^2 \exp\left(20n + \frac{5R^2}{t}\right). \end{aligned}$$

This conclues the proof of Theorem 3.1. □

3.2. **Remarks on the Optimal Log-Sobolev Constant.**

Example 3.5.

Acknowledgments. We would like to thank the Banff International Research Station (BIRS) for the opportunity to speak about the results in this paper at the Workshop [16w5025] *Analytic Versus Combinatorial in Free Probability* in December, 2016; the interactions with other workshop participants following the talk were very helpful. In particular, we thank Mireille Capitaine and Dima Shlyakhtenko for very useful conversations that led to several improvements of this work.

References

[1] Anderson, G.; Guionnet, A.; Zeitouni, O.: *An introduction to random matrices*. Vol. 118. Cambridge University Press (2010)

[2] Anderson, G.; Zeitouni, O.: *A CLT for a band matrix model.* Probab. Theory Relat. Fields **134**, 283–338 (2006)

[3] Arizmendi, O.; Nechita, I.; Vargas, C.: *On the asymptotic distribution of block-modified random matrices.* J. Math. Phys. **57** no. 1 (2016), 015216.

[4] Bai, Z.; Yin, Y.: *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix.* Ann. Probab. **16**, no. 4 (1988), 1729–1741.

[5] Bakry, D.: *L'hypercontractivité et son utilisation en thorie des semigroupes.* Lectures on probability theory (Saint-Flour, 1992), 1–114, Lecture Notes in Math., 1581, Springer, Berlin, 1994.

[6] Bakry, D.: *On Sobolev and logarithmic Sobolev inequalities for Markov semigroups.* New trends in stochastic analysis (Charingworth, 1994), 43–75, World Sci. Publ., River Edge, NJ, 1997.

[7] Bakry, D.; Barthe, F.; Cattiaux, P.; Guillin, A.: *A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case.* Electron. Commun. Probab. **13** (2008), 60–66.

[8] Bakry, D.; Émery, M.: *Diffusions hypercontractive.* Séminaire de Probabilité XIX. Lecture Notes in Math. **1123**, 179–206 (1985)

[9] Bakry, D.; Ledoux, M.: *A logarithmic Sobolev form of the Li-Yau parabolic inequality.* Rev. Mat. Iberoam. **22** no. 2, 683–702 (2006)

[10] Banica, T.; Nechita, I.: *Block-modified Wishart matrices and free Poisson laws,* Houston J. Math. **41** (2015), 113–134

[11] Banica, T.; Nechita, I.: *Asymptotic eigenvalue distributions of block-transposed Wishart matrices,* arXiv:1105.2556v2

[12] Bardet, J.; Gozlan, N.; Malrieu, F.; Zitt, P.: *Functional inequalities for Gaussian convolutions of compactly supported measures: explicit bounds and dimension dependence.* arXiv:1507.02389v1, accepted for publication in the Bernoulli Journal (2015)

[13] Bobkov, S. G.; Götze, F.: *Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities,* J. Funct. Anal. **163** (1999), 1–28.

[14] Bobkov, S.; Houdré, C.: *Some connections between isoperimetric and Sobolev-type inequalities.* Mem. Amer. Math. Soc. **129** no. 616 (1997)

[15] Bobkov, S.; Ledoux, M.: *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities.* Geom. Funct. Anal. **10**, no. 5, 1028–1052 (2000)

[16] Bobkov, S.; Tetali, P.: *Modified logarithmic Sobolev inequalities in discrete settings.* J. Theoret. Probab. **19**, no. 2, 289–336 (2006)

[17] Bogachev, S.; Molchanov, S.; Pastur, L.: *On the density of states of random band matrices.* Math. Notes **50** (1001), 1232–1242.

[18] Casati, G.; Girko, V.: *Wigner's semicircle law for band random matrices.* Random Oper. Stochastic Equations **1** (1993), 14–21.

[19] Cattiaux, P.; Guillin, A.; Wu, L.: *A note on Talagrand's transportation inequality and logarithmic Sobolev inequality.* Probab. Theory Relat. Fields **148**, 285–304 (2010)

[20] Davies, E. B.: *Explicit constants for Gaussian upper bounds on heat kernels.* Amer. J. Math. **109**, no. 2, 319–333 (1987)

[21] Davies, E. B.: *Heat kernels and spectral theory.* Cambridge Tracts in Mathematics, 92. Cambridge University Press, 1990.

[22] Davies, E. B., Simon, B.: *Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians.* J. Funct. Anal. **59**, 335–395 (1984)

[23] Diaconis, P., Saloff-Coste, L.: *Logarithmic Sobolev inequalities for finite Markov chains.* Ann. Appl. Probab. **6**, 695–750 (1996)

[24] Gross, L.: *Logarithmic Sobolev inequalities.* Amer. J. Math. **97**, 1061–1083 (1975)

[25] Guionnet, A.: *Large Random Matrices: Lectures on Macroscopic Asymptotics.* École d'Été de Probabilités de Saint-Flour XXXVI, Lecture Notes in Math., 1957, Springer, Berlin, 2009

[26] Gross, L.; Rothaus, O.: *Herbst inequalities for supercontractive semigroups.* J. Math. Kyoto Univ. **38**, no. 2, 295–318 (1998)

[27] Guionnet, A.; Zegarlinski, B.: *Lectures on logarithmic Sobolev inequalities.* Séminaire de Probabilités, XXXVI, 1134, Lecture Notes in Math., 1801, Springer, Berlin, 2003.

[28] Hoffman, A; Wielandt, H.: *The variation of the spectrum of a normal matrix.* Duke Math. J. **20**, no. 1, 37–39 (1953)

[29] Holley, R., Stroock, D.: *Logarithmic Sobolev inequalities and stochastic Ising models.* J. Statist. Phys. **46**, 1159–1194 (1987)

[30] Ledoux, M.: *Isoperimetry and Gaussian analysis. Lectures on probability theory and statistics.* 165–294, Lecture Notes in Math 1648, Springer, Berlin, 1996

[31] Ledoux, M.: *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[32] Ledoux, M.: *A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices.* Séminaire de Probabilités XXXVII, 360–369, Lecture Notes in Math., 1832, Springer, Berlin, 2003.

[33] Rashidi Far, R.; Oraby, T.; Byc, W.; Speicher, R.: *On slow-fading MIMO systems with nonseparable correlation.* IEEE Transactions on Information Thoery, Vol. **54**, No. 2, 544–553 (2008)

[34] Shlyakhtenko, D.: *Random Gaussian band matrices and freeness with amalgamation.* IMRN No. **20**, 1013–1025 (1996)

[35] Stam, A.: *Some inequalities satisfied by the quantities of information of Fisher and Shannon.* Information and Control **2**, 101–112 (1959)

[36] Villani, C.: *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003.

[37] Wigner, E.: *Characteristic Vectors of Bordered Matrices with Innite Dimensions.* Ann. of Math. **62**, 548–564 (1955)

[38] Wigner, E.: *On the Distribution of the Roots of Certain Symmetric Matrices.* Ann. of Math. **67**, 325–328 (1958)

[39] Wang, F.-Y.; Wang, J.: *Functional inequalities for convolution probability measures.* arXiv:1308.1713

[40] Yau, H.T.: *Logarithmic Sobolev inequality for the lattice gases with mixing conditions.* Commun. Math. Phys. **181**, 367–408 (1996)

[41] Yau, H.T.: *Log-Sobolev inequality for generalized simple exclusion processes.* Probab. Theory Related Fields **109**, 507–538 (1997)

[42] Zegarlinski, B.: *Dobrushin uniqueness theorem and logarithmic Sobolev inequalities.* J. Funct. Anal. **105**, 77–111 (1992)

[43] Zimmermann, D.: *Logarithmic Sobolev inequalities for mollified compactly supported measures.* J. Funct. Anal. **265**, 1064–1083 (2013)

[44] Zimmermann, D.: *Elementary proof of logarithmic Sobolev inequalities for Gaussian convolutions on* $\mathbb{R}$. arXiv:1412.1519

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, SAN DIEGO, LA JOLLA, CA 92093-0112
E-mail address: tkemp@math.ucsd.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO, 5734 S. UNIVERSITY AVENUE, CHICAGO, IL 60637
E-mail address: dz@math.uchicago.edu