ANNALES MATHÉMATIQUES

BLAISE PASCAL

Todd Kemp & David Zimmermann

Random matrices with log-range correlations, and log-Sobolev inequalities

L'accès aux articles de la revue « Annales mathématiques Blaise Pascal »
(http://ambp.centre-mersenne.org/), implique l'accord avec les conditions
générales d'utilisation (http://ambp.centre-mersenne.org/legal/).

Random matrices with log-range correlations, and log-Sobolev inequalities

Todd Kemp

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, minus its mean, converges weakly to 0 in probability; hence if the averaged empirical eigenvalue distribution converges to a law, the empirical spectral distribution converges to this limit law weakly 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 by developing a new log-Sobolev inequality which generalizes the second 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 $\mathbf{Y} + t^{1/2}\mathbf{Z}$ satisfies a log-Sobolev inequality for all $t > 0$, and we give bounds on the optimal log-Sobolev constant.

1. Introduction

Random matrix theory is primarily interested in 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 initiated by Wigner in [42, 43], 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 real Wigner ensembles as the starting point.

Given a symmetric matrix X_N, enumerate its eigenvalues $\lambda_1^N \leq \cdots \leq \lambda_N^N$ 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 proportion 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 [42, 43] 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*: real symmetric matrices with independent entries above the main diagonal, but with more complicated patterns of non-identically distributed entries (see Theorem 1.4 for a precise definition). The reader should consult 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].)

Apart from the work (of distinctly different flavor) on unitary or permutation-invariant matrix ensembles, there are comparatively few papers dealing with random matrices with *correlated entries*. In [38], 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 [3, 10, 11] 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 [37], a class of these block matrices was studied and proved to converge almost surely, with applications given to signal processing. 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 $N \to \infty$.

Remark. The actual ensembles studied in [3, 10, 11, 37, 38] 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 also that in much of this work, particularly Shlyakhtenko's

results on Gaussian band matrices [38], complex matrices were studied. In the present work, we have focused on real ensembles, but the same theorems and proofs we give here could similarly be adapted to the complex case.

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 $\{N([X_N]_{ij} - \mathbb{E}([X_N]_{ij}))^2\}_{N \in \mathbb{N}, 1 \le i, j \le N}$ 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 minus its mean converges weakly in probability to 0:

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

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 variances $\{N \text{Var}([X_N]_{ij})\}_{N \in \mathbb{N}, 1 \le i, j \le N}$ 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 minus its mean converges weakly almost surely to 0:

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

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 [3, 10, 11] but with k allowed to grow with N (with $k = o(N/(\log N)^{1/2})$), 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], which deals with independent entries.

Theorem 1.4. *Let $\{\xi_{ij}: 1 \le i \le j\}$ be zero mean unit variance i.i.d. random variables. Let $g: [0,1]^2 \to \mathbb{R}_+$ 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 [16, 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 2.10 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

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

for all sufficiently integrable positive functions f with $\int f^2 \, d\mu = 1$; here $\text{Ent}_\mu(g) = \int g \log g \, d\mu$ for a μ-probability density g. The inequality (1.4) first appeared in [39] (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 [26] 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 in probability theory, functional analysis, and differential geometry; see, for example, [5, 6, 13, 15, 20, 21, 22, 23, 29, 33, 34, 35, 40, 44, 45, 46].

There is a big industry of literature devoted to necessary and sufficient conditions for a log-Sobolev inequality to hold; cf. [8, 9, 14, 27, 31]. Adding to these efforts (with applications to random matrices in mind), the second author of the present paper developed a new approximation scheme, the *mollified log-Sobolev inequality*, in [47]: if Y is any bounded random variable and Z is a standard normal random variable independent from Y, then the law of $Y + t^{1/2}Z$ satisfies a log-Sobolev inequality for all $t > 0$, with a constant $c(t)$ that is bounded in terms of an exponential of $\|Y - \mathbb{E}[Y]\|_\infty^2/t$. The following result is a generalization of those one-dimensional ideas to higher dimension.

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. $\text{Law}_{\mathbf{Z}}(d\mathbf{x}) = (2\pi)^{-d/2}e^{-|\mathbf{x}|^2/2}\,d\mathbf{x}$) independent from $\mathbf{Y}$. For $0 < t \leq \|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}$, the measure $\text{Law}_{\mathbf{Y}+t^{1/2}\mathbf{Z}}$ satisfies a log-Sobolev inequality, with constant $c(t)$ satisfying*

$$c(t) \leq 289\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}^2 \exp\left(20d + \frac{5\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}^2}{t}\right). \tag{1.5}$$

Remark 1.7. Logarithmic Sobolev inequalities are invariant under translations of the measure (as is evident from (1.4)). Since $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$, we could just as well give the bound (1.5) on $c(t)$ with $\mathbf{Y} - \mathbb{E}[\mathbf{Y}]$ replaced simply by $\mathbf{Y}$. In most applications it is better to work with the centered vector $\mathbf{Y} - \mathbb{E}[\mathbf{Y}]$, but it will be useful in the proof of the second statement of Theorem 1.1 (regarding bounded random matrix ensembles) to use it in the un-centered form.

We briefly expound the history of Theorem 1.6. Following the second author's paper [47], in [41] 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.

Further complicating this history: an early version of the present paper, posted on the arXiv, proved Theorem 1.6 as its central result. In response, Bardet, Gozlan, Malrieu, and Zitt [12], building on our techniques, sharpened the inequality (1.5) to a stronger form with linear, rather than exponential, dependence on dimension:

$$c(t) \leq \left(K_1 d + K_2 \frac{\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}^2}{t} \right) \|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}^2 \exp\left(\frac{4\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}^2}{t} \right) \qquad (1.6)$$

for some universal constants K_1, K_2. Our proof of (1.5) relies on an estimate for the best constant in the Poincaré inequality, which we were only able to prove with a dimension-dependent bound. The main contribution to this problem in [12] was a dimension-independent bound on the Poincaré constant, [12, Theorems 1.2–1.3].

Remark 1.8. We do not know if the optimal constant $c(t)$ in (1.5) 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 $\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\|_{L^\infty}$ that forces the blocks to be of size $o(\log N)$; and this dependence is sharp, as was shown in the second author's paper [47, Theorem 15].

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 2.10) 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 2.10, 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

The connection between random matrices and log-Sobolev inequalities that we exploit in this paper was introduced by Guionnet in [28]. Using the Herbst inequality [27], which yields Gaussian concentration of Lipschitz functionals, she offered a fundamentally new proof of Wigner's semicircle law; this proof automatically generalized to non-Gaussian ensembles whose entries satisfy a log-Sobolev inequality. The paper [47] was motivated by generalizing this method, to give a proof technique of Wigner's law that would apply (by approximation) to ensembles whose entries do not satisfy a log-Sobolev inequality. That approach, working entry-by-entry, required the entries to be independent. The main goal of this paper is to weaken that assumption, and so we present here a brief discussion of Guionnet's approach, with an eye towards removing independence assumptions.

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 $\lambda_1^N \leq \cdots \leq \lambda_N^N$, and let μ_N denote the empirical spectral distribution (ESD) of X_N; cf. (1.1). Wigner's law [42, 43] 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 [28, p. 70, Theorem 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 $f : \mathbb{R} \to \mathbb{R}$,*

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

In fact, in the Wigner ensemble 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 [26, p. 1074, Remark 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 $\nu_1 \otimes \nu_2$ 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 [47], the second author showed that, under certain integrability conditions, the empirical law of eigenvalues μ_N minus its mean converges weakly in probability to 0, *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 plus a GOE_N noise of 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 in terms of the *joint* distribution in Theorem 2.1. Guionnet proved the theorem from the Herbst concentration inequality [27], 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 functional calculus, together with the following lemma from matrix theory (see [30, p. 37, Theorem 1, and p. 39, Remark 2]).

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

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

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, $\|\sqrt{N}[X_N]_{ij}\|_\infty \leq R$ for all N and all $1 \leq i, j \leq N$. (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

$$\widetilde{X}_N = X_N + t^{1/2}G_N. \tag{2.1}$$

For $1 \leq k \leq r$, let $\mathbf{Y}_k$ denote the random vector in $\mathbb{R}^d = \mathbb{R}^{|P_k|}$ 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 $\leq R$, and so $\|\sqrt{N}\mathbf{Y}_k\|_{L^\infty} \leq R|P_k|^{1/2} \leq Rd_N^{1/2}$. 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 (cf. Remark 1.7)

$$c(t) \leq 289R^2 d_N \exp\left(20|P_k| + \frac{5 \cdot R^2 d_N}{t}\right).$$

Let us (wlog) assume that $R \geq 1$ and $t \leq 1$. Then $|P_k| \leq d_N \leq R^2 d_N/t$, and also $R^2 d_N \leq R^2 d_N/t \leq \exp(R^2 d_N/t)$. Thus

$$c(t) \leq 289R^2 d_N \exp\left(20d_N + \frac{5R^2 d_N}{t}\right) \leq 289\exp\left(\frac{21R^2 d_N}{t}\right). \tag{2.2}$$

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 $\widetilde{X}_N$. Let μ_N denote the ESD of $X_N - \mathbb{E}[X_N]$, and let $\widetilde{\mu}_N$ denote the ESD of the mollified version $\widetilde{X}_N$. Then, for each $\epsilon > 0$, and each test function f, we have the following standard triangle inequality estimate.

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

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

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

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 $f \in \mathrm{Lip}(\mathbb{R})$,*

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

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

$$\left| \int f \, d\mu_N - \int f \, d\widetilde{\mu}_N \right| = \frac{1}{N} \left| \sum_{j=1}^{N} [f(\lambda_j^N) - f(\widetilde{\lambda}_j^N)] \right| \leq \frac{1}{N} \sum_{i=1}^{N} \|f\|_{\mathrm{Lip}} \left| \lambda_i^N - \widetilde{\lambda}_i^N \right|$$

$$\leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}} \left(\sum_{i=1}^{N} (\lambda_i^N - \widetilde{\lambda}_i^N)^2 \right)^{1/2}$$

$$\leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}} \left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2] \right)^{1/2}.$$

Now, for any symmetric $N \times N$ matrix A,

$$\left(\frac{1}{N} \mathrm{Tr}(A^2) \right)^{1/2} \leq \|A\|_{\mathrm{op}}$$

(indeed, this is nothing more than the usual inequality between ℓ^2 and ℓ^∞ norms on $\mathbb{R}^N$, applied to the eigenvalues of A). Applying this to $A = X_N - \widetilde{X}_N = t_N^{1/2} G_N$, we therefore have

$$\left| \int f \, d\mu_N - \int f \, d\widetilde{\mu}_N \right| \leq \|f\|_{\mathrm{Lip}} \|X_N - \widetilde{X}_N\|_{\mathrm{op}} = t_N^{1/2} \|f\|_{\mathrm{Lip}} \|G_N\|_{\mathrm{op}} \text{ a.s.}$$

According to [4], $\limsup_{N \to \infty} \|G_N\|_{\mathrm{op}} \leq 2$ a.s. This proves the result. $\qquad \square$

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

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

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

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

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 $\leq \frac{1}{N^2}$. The result now follows from the Borel–Cantelli lemma. $\qquad \square$

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

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

Proof. We simply follow calculations like the ones in the proof of Lemma 2.5:

$$\mathbb{E}\left(\left| \int f \, d\widetilde{\mu}_N - \int f \, d\mu_N \right| \right) \leq \mathbb{E}\left(\frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}} \left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2] \right)^{1/2} \right)$$

$$\leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}} \left(\mathbb{E}\left(\mathrm{Tr}[(X_N - \widetilde{X}_N)^2] \right) \right)^{1/2} = \|f\|_{\mathrm{Lip}} t_N^{1/2}$$

where we applied Jensen's inequality in the second step. Since $t_N \to 0$, the result follows. $\qquad \square$

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 $c(t_N) = o\left(\frac{N^2}{\log N}\right)$. For N sufficiently large, we define

$$t_N := 21R^2 \frac{d_N}{\log N}.$$

By Assumption (2) of Theorem 1.1, $d_N = o(\log N)$, and hence $t_N \to 0$ as $N \to \infty$. From (2.2), we have

$$c(t_N) \leq 289 \exp\left(\frac{21R^2 d_N}{21R^2 d_N / \log N} \right) = 289N = o\left(\frac{N^2}{\log N} \right).$$

This concludes the proof. $\qquad \square$

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 we had made use of the stronger (1.6) from [12], we could not have avoided a factor of $e^{Kd_N/t}$ (for some constant K) in the estimate for the log-Sobolev constant $c(t)$, as the reader can quickly verify. In fact, in [47], the second author showed that the optimal log Sobolev inequality (in one dimension) typically has this $e^{K/t}$ form. As such, the assumption that $d_N = o(\log N)$ cannot be weakened, and 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 $R \geq 0$ such that

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

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

$$[\widehat{X}_N]_{ij} = [X_N]_{ij} \cdot \mathbb{1}_{\{\sqrt{N}|([X_N]_{ij} - \mathbb{E}([X_N]_{ij}))|\leq R\}}.$$

Then $\|\sqrt{N}\widehat{X}_{ij}\|_\infty \leq R$ for all N, i, j. Let $\widehat{\mu}_N$ denote the ESD of $\widehat{X}_N$. The preceding proof shows that $\int f \, d\widehat{\mu}_N$ minus its mean converge to 0 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:

$$\mathbb{P}\left(\left|\int f \, d\mu_N - \mathbb{E}\left(\int f \, d\mu_N\right)\right| \geq \epsilon\right)$$

$$\leq \mathbb{P}\left(\left|\int f \, d\mu_N - \int f \, d\widehat{\mu}_N\right| \geq \frac{\epsilon}{3}\right) + \mathbb{P}\left(\left|\int f \, d\widehat{\mu}_N - \mathbb{E}\left(\int f \, d\widehat{\mu}_N\right)\right| \geq \frac{\epsilon}{3}\right)$$

$$+ \mathbb{P}\left(\left|\mathbb{E}\left(\int f \, d\widehat{\mu}_N\right) - \mathbb{E}\left(\int f \, d\mu_N\right)\right| \geq \frac{\epsilon}{3}\right). \quad (2.6)$$

The above proof in the uniformly 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:

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

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

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

$$\left|\mathbb{E}\left(\int f \, d\widehat{\mu}_N\right) - \mathbb{E}\left(\int f \, d\mu_N\right)\right| \leq \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}\left(\mathbb{E}\left(\mathrm{Tr}[(X_N - \widehat{X}_N)^2]\right)\right)^{1/2}$$

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

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.

2.3. Theorem 1.2, a Generalization, and Applications

We begin with a lemma which appeared in the second author's paper [48, Proposition 6], but was surely folklore far earlier. We reproduce the simple proof here, for completeness.

Lemma 2.9. *Let* $\mathbf{X}$ *be a random vector in* $\mathbb{R}^d$ *whose law satisfies a log-Sobolev inequality* (1.4) *with constant c. Let* $T : \mathbb{R}^d \to \mathbb{R}^d$ *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 $f : \mathbb{R}^d \to \mathbb{R}$ be a non-negative locally Lipschitz function. Then $f \circ T$ is locally-Lipschitz and non-negative. Since μ satisfies the LSI with constant c, it follows that

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

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. $\qquad\square$

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

Let $X_N = T_N(N^{-1/2}\Xi_N)$, *and let* μ_N *denote the ESD of* X_N. *Then* μ_N *minus its mean converges to* 0 *almost surely:*

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

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 2.9, $S_N(\Xi_N) = \sqrt{N}X_N$ 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 $f \in \mathrm{Lip}(\mathbb{R})$,

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

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

We now prove Theorem 1.2, as a Corollary to Theorem 2.10.

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. In the special case that T is invertible, this is equivalent to $\mathbf{X} - \mathbb{E}(\mathbf{X}) = A\mathbf{G}$ for some invertible matrix A. Since $A\mathbf{G}$ has a density equal to a constant times $\exp(-\frac{1}{2}\mathbf{x} \cdot (AA^*)^{-1}\mathbf{x})$, this yields the more standard definition of a (non-degenerate) "jointly Gaussian" random vector.

Let $\Pi_N = \{P_1, \ldots, P_r\}$ denote the partition of $\{(i,j)\colon 1 \leq i \leq j \leq N\}$ in the theorem, and for $1 \leq k \leq r$ 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 $\mathbf{X}_k = T^k(N^{-1/2}\mathbf{G}_k)$, where $\mathbf{G}_k$ is a standard Gaussian random vector in $\mathbb{R}^{|P_k|}$. The entries of all the $\mathbf{G}_k$ are i.i.d. standard normal random variables, which each satisfy log-Sobolev inequalities with constant 1 (cf. [26]). Hence, letting $T_N \colon \mathbb{M}_N^{\mathrm{sym}} \to \mathbb{M}_N^{\mathrm{sym}}$ be the map which takes the entries in partition block P_k to T^k of those entries, we see that T_N is a Lipschitz function, with Lipschitz norm equal to $\max\{\|T^k\|_{\mathrm{Lip}}, 1 \leq k \leq r\}$.

We proceed to estimate the Lipschitz norm of the affine map T^k. This is just the operator norm of its linear part $\mathring{T}^k = T^k - N^{1/2}\mathbb{E}(\mathbf{X}_k)$. The operator norm is bounded above by the Hilbert–Schmidt norm; thus $\|T^k\|_{\mathrm{Lip}} \leq \|\mathring{T}^k\|_{\mathrm{HS}}$. Note that

$$\|\mathring{T}^k\|_{\mathrm{HS}}^2 = \sum_{a,b=1}^{|P_k|} [\mathring{T}^k]_{ab}^2$$

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

$$\|\hat{T}^k\|_{\text{HS}}^2 = \sum_{a=1}^{|P_k|} \text{Var}(N^{1/2}[\mathbf{X}_k]_a) \leq R^2|P_k| \leq R^2 d_N.$$

By assumption, $d_N = o\left(\frac{N^2}{\log N}\right)$, and so we have shown that $\|T_N\|_{\text{Lip}} = o\left(\sqrt{\frac{N^2}{\log N}}\right)$. The result now follows from Theorem 2.10. $\qquad\square$

2.4. Theorem 1.4, Generalizations, and Examples

In this section, we show how to prove Theorem 1.4 as a straightforward corollary to Theorems 1.1 and 2.10. To begin, we note that the topic of the paper [38] 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. [38, 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 centered ESD to 0 in probability, it suffices to show that the family $\{N[X_N]_{ij}^2\}_{N \in \mathbb{N}, 1 \leq i, j \leq N}$ 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 $\mathbb{E}(N[X_N]_{ij}^2) \leq \|g\|_\infty \mathbb{E}(\xi_{ij}^2) = \|g\|_\infty$ 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 centered ESD of X_N converges to 0 in probability.

For the second statement of the theorem: first, if ξ_{ij} are bounded random variables $\leq R$, then the entries $\sqrt{N}|[X_N]_{ij}| \leq \|g\|_\infty R$ 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 2.10. 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 2.10 to show that $\|T_N\|_{\text{Lip}} = o\left(\frac{N}{\sqrt{\log N}}\right)$. But since T_N is linear and diagonal, its Lipschitz norm (i.e. operator norm) is simply the maximum modulus of the entries, $\|T_N\|_{\text{op}} = \max_{i,j}|g(i/N, j/N)| \le \|g\|_\infty = O(1) = o\left(\frac{N}{\sqrt{\log N}}\right)$. This concludes the proof. □

To conclude, let's consider a large family of examples to which the above results apply, this time with correlated entries.

Example 2.11. In [17], the authors analyzed the random Toeplitz matrix: an $N \times N$ random matrix ensemble with independent *diagonals* but equal entries along each diagonal. That is: $[X_N]_{ij} = X_{|i-j|}$ where $\{X_0, \ldots, X_{N-1}\}$ are i.i.d. random variables. They showed, under appropriate rescaling, that the empirical spectral distribution of this ensemble converges to a heretofore unknown unbounded probability distribution, now known as the Toeplitz law. Their methods, which were both combinatorial and probabilistic, showed both universality (that the resultant bulk spectral distribution is independent of the common law of the entries $X_0, \ldots, X_{N-1}$ provided there are two finite moments), and that the convergence is almost sure. To the second point: they use the method of moments to prove convergence in expectation, and then leverage the precise structure of the correlated blocks (i.e. diagonals) to estimate the rate of convergence tightly enough to upgrade to a.s. convergence.

There have been a number of generalizations of these results in recent years: namely in [36], (which studied circulant, i.e. periodic, variations on the Toeplitz ensemble), [32] (which considered a banded version of the random Toeplitz ensemble), and [25] (which studied the more general case where the entries along each diagonal are not equal, but each pair has the same correlation coefficient). In [25, 32], only convergence in expectation was proved. Since the correlated blocks in these ensembles (the diagonals) are all of size $O(N) = o\left(\frac{N^2}{\log N}\right)$, our Theorem 1.2 implies a.s. convergence of the empirical spectral distributions, in the case that the entries are Gaussian (or satisfy the more general Lipschitz constraints of Theorem 2.10).

We can consider a more general family of matrix matrices with independent diagonals, and entries along each diagonal with a given covariance structure that provides some hope of a large-N limit for the bulk eigenvalue distribution. This is the topic of a forthcoming preprint [24] by the present first author and his student coauthors. All such ensembles
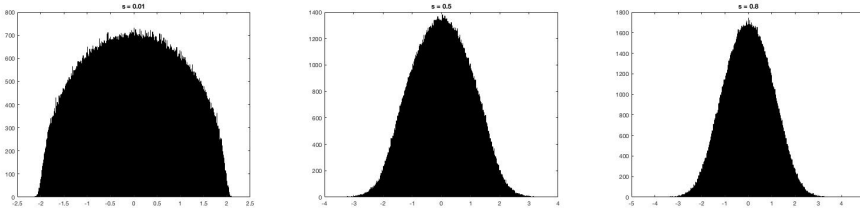
FIGURE 2.1. Histograms of the empirical spectral distribution of the s-band ensemble for $s = 0.01, 0.5, 0.8$. These simulations were performed with 1000 trials at $N = 200$.

have correlated blocks of size $O(N)$, and hence Theorem 1.2 or 2.10 yield automatic upgrades from convergence in expectation to a.s. convergence of the empirical spectral distribution. One good example of such models where convergence in expectation can be established by combinatorial means is the following kind of mixture of the models in [17] and [32]: let b_N be a "band-width", $1 \leq b_N \leq N$, and consider the ensemble X_N where, for $|i - j| \leq b_N$, $[X_N]_{ij}$ follow the random Toeplitz ensemble (i.e. $[X_N]_{ij} = X_{|i-j|}$ for some i.i.d. random variables $X_0, X_1, X_2, \ldots$), while for $|i - j| > b_N$, $[X_N]_{ij}$ are i.i.d. for $i < j$, with the same distribution as the entries in the band. If $\lim_{N \to \infty} \frac{b_N}{N} = s \in (0, 1)$, the empirical spectral distribution of this ensemble converges in expectation to a law γ_s which is unbounded, and different for different values of s. As $s \to 1$, we recover the Toeplitz law, while for $s \to 0$ we recover the semicircle law. See Figure 2.1 for simulations. Again, since the correlated blocks are all of size $O(N)$, by Theorem 1.2 or 2.10, these "s-band" ensembles all converge a.s. when the entries are Gaussian.

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 $0 < t \leq R^2$, i.e., $\gamma_t(x) = (2\pi t)^{-d/2} \exp(-\frac{|x|^2}{2t}) \, dx$. Then the convolution $\mu * \gamma_t$ satisfies the log-Sobolev inequality with optimal constant $c(t)$ bounded by*

$$c(t) \leq 289 \, R^2 \exp\left(20d + \frac{5R^2}{t}\right).$$

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

Theorem 3.2 (Cattiaux, Guillin, Wu). *Let μ be a probability measure on $\mathbb{R}^d$ with $\mathrm{d}\mu(x) = e^{-V(x)}\mathrm{d}x$ for some $V \in C^2(\mathbb{R}^d)$. Suppose the following:*

(1) *There exists a constant $K \leq 0$ such that $\mathrm{Hess}(V) \geq KI$.*

(2) *There exists a $W \in C^2(\mathbb{R}^d)$ with $W \geq 1$ and constants $b, c > 0$ such that*
$$\Delta W(x) - \langle \nabla V, \nabla W \rangle(x) \leq (b - c|x|^2)W(x)$$
for all $x \in \mathbb{R}^d$.

Then μ satisfies a LSI.

In particular, let $r_0, b', \lambda > 0$ be such that
$$\Delta W(x) - \langle \nabla V, \nabla W \rangle(x) \leq -\lambda W(x) + b'\mathbb{1}_{B_{r_0}}$$
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, Theorem. 1.4]*, μ satisfies a Poincaré inequality with some constant C_P; that is, for every sufficiently smooth g with $\int g\,\mathrm{d}\mu = 0$,*

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

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

$$\kappa_{r_0} \leq \frac{4}{\pi^2} r_0^2 \frac{\sup_{x \in B_{r_0}} p(x)}{\inf_{x \in B_{r_0}} p(x)},$$

where $p(x) = e^{-V(x)}$ is the density of μ. Let

$$A = \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right) + \epsilon$$

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

where $\epsilon > 0$ is an arbitrarily chosen parameter. Then μ satisfies a LSI with constant $A + (B + 2)C_P$.

We remark that the statement of Theorem 3.2 is a combination of results in [19] and preceding papers cited therein (notably [7]). In those papers, they are proved in the more general context of Riemannian manifolds. Also, the constants given above are derived in the proofs in [19] but not presented in the concise form above. The reader is directed to [41, pp. 7–8] for the precise statement above.

With the above, we now prove Theorem 3.1.

Proof of Theorem 3.1. By translation invariance of LSI, we will assume that μ is supported in B_R. We will apply Theorem 3.2 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 [41, pp. 7–8]. Let $Q_t(x) = \frac{x^2}{2t} + \frac{d}{2} \log(2\pi t)$, let p_t be the density of μ_t (which is smooth and strictly positive, since $\mu_t = \mu * \gamma_t$), and set $V_t(x) = -\log(p_t(x))$. Thus

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

Also let

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

so ν_x is a probability measure for each $x \in \mathbb{R}^d$. For any smooth function F of $x \in \mathbb{R}^d$ and any vector $X \in \mathbb{R}^d$ with $|X| = 1$, let $\nabla_X F(x)$ denote the directional derivative of F at x in the direction X, and let $\text{Hess}(F)(X, X)(x) := \nabla_X^2 F(x)$ denote the second directional derivative of F in the direction X at the point x. Following the proof of [41, Theorem 2.6 (a)] (i.e. using the chain rule applied to the first and second derivatives of e^{-Q_t} and e^{-V_t}), we see the relevance of the measure ν_x:

$\text{Hess}(V_t)(X, X)(x)$

$$= \left| \int_{B_R} \nabla_X Q_t(x-z) d\nu_x(z) \right|^2 - \int_{B_R} \left(|\nabla_X Q_t(x-z)|^2 - \text{Hess}(Q_t)(X, X)(x-z) \right) d\nu_x(z)$$

$$= \frac{1}{t} - \left(\int_{B_R} |\nabla_X Q_t(x-z)|^2 d\nu_x(z) - \left| \int_{B_R} \nabla_X Q_t(x-z) d\nu_x(z) \right|^2 \right),$$

since $\text{Hess}(Q_t) = \frac{1}{t}$. But for any C^1 function f,

$$\int_{B_R} f^2 d\nu_x(z) - \left(\int_{B_R} f \, d\nu_x(z) \right)^2 = \frac{1}{2} \int_{B_R \times B_R} (f(z) - f(y))^2 d\nu_x(z) d\nu_x(y)$$

$$\leq 2R^2 \sup |\nabla f|^2,$$

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

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

So we take $K = \frac{1}{t} - \frac{2R^2}{t^2}$. Note that $K \leq 0$ since $t \leq R^2$. Let

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

Then

$$\frac{\Delta W - \langle \nabla V_t, \nabla W \rangle}{W}(x) = \frac{d}{8t} + \frac{|x|^2}{64t^2} - \frac{1}{16t} \int_{B_R} \langle x, \nabla Q_t(x-z) \rangle \mathrm{d}\nu_x(z)$$

$$= \frac{d}{8t} + \frac{|x|^2}{64t^2} - \frac{1}{16t^2} \int_{B_R} \left(|x|^2 - \langle x, z \rangle \right) \mathrm{d}\nu_x(z)$$

$$\leq \frac{d}{8t} - \frac{3|x|^2}{64t^2} + \frac{1}{16t^2} \sup_{z \in B_R} \langle x, z \rangle$$

$$= \frac{d}{8t} - \frac{3|x|^2}{64t^2} + \frac{1}{16t^2} R|x|.$$

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

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

so we take

$$b = \frac{d}{8t} + \frac{R^2}{32t^2},$$

$$c = \frac{1}{64t^2}.$$

Now let

$$r_0 = \sqrt{16dt + 2R^2},$$

$$b' = \frac{1}{4t} \exp \left(d + \frac{R^2}{8t} - 1 \right),$$

$$\lambda = \frac{d}{8t}.$$

We claim that

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

so that

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

We have

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

$$= \left(d + \frac{R^2}{8t} - \frac{|x|^2}{16t} \right) \exp \left(- \left(d + \frac{R^2}{8t} - \frac{|x|^2}{16t} \right) + 1 \right).$$

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

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

and

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

$$= (2\pi t)^{-d/2} \exp\left(-\frac{(r_0+R)^2}{2t}\right).$$

Thus

$$\kappa_{r_0} \leq \frac{4}{\pi^2} r_0^2 \frac{\sup_{x \in B_{r_0}} p(x)}{\inf_{x \in B_{r_0}} p(x)} \leq D r_0^2 \exp\left(\frac{(r_0+R)^2}{2t}\right).$$

Now, following (3.1) in Theorem 3.2, we take

$$C_P = \frac{1 + b' \kappa_{r_0}}{\lambda}$$

$$\leq \frac{8t}{d}\left(1 + \frac{1}{4t} \exp\left(d + \frac{R^2}{8t} - 1\right) \cdot \frac{4}{\pi^2} r_0^2 \exp\left(\frac{(r_0+R)^2}{2t}\right)\right)$$

$$= \frac{8t}{d} + \frac{4}{\pi^2 e}\left(32t + \frac{4R^2}{d}\right) \exp\left(d + \frac{R^2}{8t} + \frac{(\sqrt{16dt + 2R^2} + R)^2}{2t}\right).$$

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

$$C_P \leq \frac{8R^2}{1} + \frac{4}{\pi^2 e}\left(32R^2 + \frac{4R^2}{1}\right) \exp\left(d + \frac{R^2}{8t} + \frac{\sqrt{2(16dt + 2R^2 + R^2)}^2}{2t}\right)$$

$$= 8R^2 + \frac{144}{\pi^2 e} R^2 \exp\left(17d + \frac{25R^2}{8t}\right)$$

$$\leq \left(8 + \frac{144}{\pi^2 e}\right) R^2 \exp\left(17d + \frac{25R^2}{8t}\right).$$

Next, we estimate $\int |x|^2 d\mu_t(x)$. If Z is a random vector with law μ and G is a standard normal random vector in $\mathbb{R}^d$ independent from Z, then μ_t is the law of $Z + t^{1/2}G$. Thus by the change of variables theorem,

$$\int_{\mathbb{R}^d} |x|^2 d\mu_t(x) = \mathbb{E}\left[|Z + t^{1/2}G|^2\right] = \mathbb{E}[|Z|^2] + 2t^{1/2}\mathbb{E}[Z \cdot G] + t\mathbb{E}[|G|^2].$$

Since supp $\mu \subseteq B_R$, $|Z| \le R$ a.s. Also, by the independence of Z and G, $\mathbb{E}[Z \cdot G] = 0$. Finally, each component of G is a standard normal with second moment 1. Thus, we have the estimate

$$\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu_t(x) \le td + R^2.$$

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

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

and

$$B = \frac{2}{c}\left(\frac{1}{\epsilon} - \frac{K}{2}\right)\left(b + c\int |x|^2 \mathrm{d}\mu_t(x)\right)$$

$$\le 128t^2\left(\frac{1}{16t} - \left(\frac{1}{2t} - \frac{R^2}{t^2}\right)\right)\left(\frac{d}{8t} + \frac{R^2}{32t^2} + \frac{1}{64t^2}\left(td + R^2\right)\right)$$

$$= \frac{18dR^2}{t} + \frac{6R^4}{t^2} - \frac{63d}{8} - \frac{21R^2}{8}$$

$$\le \frac{18dR^2}{t} + \frac{6R^4}{t^2} - 2.$$

Putting everything together, applying the conclusion of Theorem 3.2, we get that the optimal log-Sobolev constant $c(t)$ for μ_t satisfies

$$c(t) \le A + (B + 2)C_P$$

$$\le 128R^2 + \left(\frac{18dR^2}{t} + \frac{6R^4}{t^2} - 2 + 2\right)\left(8 + \frac{144}{\pi^2 e}\right)R^2\exp\left(17d + \frac{25R^2}{8t}\right)$$

$$= 128R^2 + 12 \cdot \frac{R^2}{2t}\left(3d + \frac{R^2}{t}\right)\left(8 + \frac{144}{\pi^2 e}\right)R^2\exp\left(17d + \frac{25R^2}{8t}\right).$$

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

$$c(t) \le 128R^2 + 12\exp\left(\frac{R^2}{2t}\right)\exp\left(3d + \frac{R^2}{t}\right)\left(8 + \frac{144}{\pi^2 e}\right)R^2\exp\left(17d + \frac{25R^2}{8t}\right)$$

$$= 128R^2 + \left(96 + \frac{1728}{\pi^2 e}\right)R^2\exp\left(20d + \frac{37R^2}{8t}\right)$$

$$\le \left(128 + 96 + \frac{1728}{\pi^2 e}\right)R^2\exp\left(20d + \frac{5R^2}{t}\right)$$

$$< 289R^2\exp\left(20d + \frac{5R^2}{t}\right).$$

This concludes the proof of Theorem 3.1. $\qquad\qquad\square$

Acknowledgments

References

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2010.

[2] Greg W. Anderson and Ofer Zeitouni. A CLT for a band matrix model. *Probab. Theory Relat. Fields*, 134(2):283–338, 2006.

[3] Octavio Arizmendi, Ion Nechita, and Carlos Vargas. On the asymptotic distribution of block-modified random matrices. *J. Math. Phys.*, 57(1): article no. 015216 (25 pages), 2016.

[4] Zhi-Dong Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16(4):1729–1741, 1988.

[5] Dominique Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Mathematics*, pages 1–114. Springer, 1994.

[6] Dominique Bakry. On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. In *New trends in stochastic analysis (Charingworth, 1994)*, pages 43–75. World Scientific, 1997.

[7] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.*, 13:60–66, 2008.

[8] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, 1985.

[9] Dominique Bakry and Michel Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.*, 22(2):683–702, 2006.

[10] Teodor Banica and Ion Nechita. Asymptotic eigenvalue distributions of block-transposed Wishart matrices. *J. Theor. Probab.*, 26(3):855–869, 2013.

[11] Teodor Banica and Ion Nechita. Block-modified Wishart matrices and free Poisson laws. *Houston J. Math.*, 41(1):113–134, 2015.

[12] Jean-Baptiste Bardet, Nathaël Gozlan, Florent Malrieu, and Pierre-André Zitt. Functional inequalities for Gaussian convolutions of compactly supported measures: explicit bounds and dimension dependence. *Bernoulli*, 24(1):333–353, 2018.

[13] Sergey G. Bobkov and Christian Houdré. Some connections between isoperimetric and Sobolev-type inequalities. *Mem. Am. Math. Soc.*, 129(616):viii+111, 1997.

[14] Sergey G. Bobkov and Michel Ledoux. From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[15] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theor. Probab.*, 19(2):289–336, 2006.

[16] Leonid V. Bogachev, Stanislav A. Molchanov, and Leonid A. Pastur. On the density of states of random band matrices. *Mat. Zametki*, 50(6):31–42, 1991.

[17] Włodzimierz Bryc, Amir Dembo, and Tiefeng Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):1–38, 2006.

[18] Giulio Casati and Vyacheslav Girko. Wigner's semicircle law for band random matrices. *Random Oper. Stoch. Equ.*, 1(1):15–21, 1993.

[19] Patrick Cattiaux, Arnaud Guillin, and Li-Ming Wu. A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Relat. Fields*, 148(1-2):285–304, 2010.

[20] Edward B. Davies. Explicit constants for Gaussian upper bounds on heat kernels. *Am. J. Math.*, 109(2):319–333, 1987.

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

[22] Edward B. Davies and Barry Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.*, 59(2):335–395, 1984.

[23] Persi Diaconis and Laurent Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[24] R. Elliott Smith, H. Huang, T. Kemp, Y. Ling, X. Luo, E. Lybrand, and J. Wang. Random Matrices with Independent Diagonals. in preparation.

[25] Olga Friesen and Matthias Löwe. A phase transition for the limiting spectral density of random matrices. *Electron. J. Probab.*, 18: article no. 17 (17 pages), 2013.

[26] Leonard Gross. Logarithmic Sobolev inequalities. *Am. J. Math.*, 97(4):1061–1083, 1975.

[27] Leonard Gross and Oscar Rothaus. Herbst inequalities for supercontractive semi-groups. *J. Math. Kyoto Univ.*, 38(2):295–318, 1998.

[28] Alice Guionnet. *Large random matrices: lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[29] Alice Guionnet and Bogusław Zegarliński. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Mathematics*, pages 1–134. Springer, 2003.

[30] Alan J. Hoffman and Helmut W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Math. J.*, 20:37–39, 1953.

[31] Richard Holley and Daniel Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.*, 46(5-6):1159–1194, 1987.

[32] Vladislav Kargin. Spectrum of random Toeplitz matrices with band structure. *Electron. Commun. Probab.*, 14:412–421, 2009.

[33] Michel Ledoux. Isoperimetry and Gaussian analysis. In *Lectures on probability theory and statistics (Saint-Flour, 1994)*, volume 1648 of *Lecture Notes in Mathematics*, pages 165–294. Springer, 1996.

[34] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[35] Michel Ledoux. A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices. In *Séminaire de Probabilités XXXVII*, volume 1832 of *Lecture Notes in Mathematics*, pages 360–369. Springer, 2003.

[36] Adam Massey, Steven J. Miller, and John Sinsheimer. Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices. *J. Theor. Probab.*, 20(3):637–662, 2007.

[37] Reza Rashidi Far, Tamer Oraby, Wlodek Bryc, and Roland Speicher. On slow-fading MIMO systems with nonseparable correlation. *IEEE Trans. Inf. Theory*, 54(2):544–553, 2008.

[38] Dimitri Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *Int. Math. Res. Not.*, 1996(20):1013–1025, 1996.

[39] Aart J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inform. and Control*, 2:101–112, 1959.

[40] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.

[41] Feng-Yu Wang and Jian Wang. Functional inequalities for convolution probability measures. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 52(2):898–914, 2016.

[42] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.*, 62:548–564, 1955.

[43] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. Math.*, 67:325–327, 1958.

[44] Horng-Tzer Yau. Logarithmic Sobolev inequality for lattice gases with mixing conditions. *Commun. Math. Phys.*, 181(2):367–408, 1996.

[45] Horng-Tzer Yau. Logarithmic Sobolev inequality for generalized simple exclusion processes. *Probab. Theory Relat. Fields*, 109(4):507–538, 1997.

[46] Bogusław Zegarliński. Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 105(1):77–111, 1992.

[47] David Zimmermann. Logarithmic Sobolev inequalities for mollified compactly supported measures. *J. Funct. Anal.*, 265(6):1064–1083, 2013.

[48] David Zimmermann. Elementary proof of logarithmic Sobolev inequalities for Gaussian convolutions on $\mathbb{R}$. *Ann. Math. Blaise Pascal*, 23(1):129–140, 2016.

TODD KEMP
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
USA
tkemp@ucsd.edu

DAVID ZIMMERMANN
davidszimmermann@gmail.com