Journal of Statistical Physics, Vol. 105, Nos. 1/2, October 2001 (© 2001)

Almost Sure Weak Convergence for the Generalized Orthogonal Ensemble

Gordon Blower[1]

Received January 5, 2001; revised May 23, 2001

The generalized orthogonal ensemble satisfies isoperimetric inequalities analogous to the Gaussian isoperimetric inequality, and an analogue of Wigner's law. Let v be a continuous and even real function such that $V(X) = \text{trace } v(X)/n$ defines a uniformly p-convex function on the real symmetric $n \times n$ matrices X for some $p \geqslant 2$. Then $\nu(dX) = e^{-V(X)} dX/Z$ satisfies deviation and transportation inequalities analogous to those satisfied by Gaussian measure[6,27], but for the Schatten c^p norm. The map, that associates to each $X \in M_n^s(\mathbb{R})$ its ordered eigenvalue sequence, induces from ν a measure which satisfies similar inequalities. It follows from such concentration inequalities that the empirical distribution of eigenvalues converges weakly almost surely to some non-random compactly supported probability distribution as $n \to \infty$.

KEY WORDS: Random matrices; transportation; isoperimetric inequality; statistical mechanics.

1. INTRODUCTION AND MAIN RESULTS

We consider an ensemble of real symmetric $n \times n$ random matrices which generalizes the Gaussian orthogonal ensemble of Dyson and Mehta[20] and arises in quantum gravity. We show that the empirical distribution of eigenvalues converges weakly almost surely as $n \to \infty$, thereby providing an analogue of Wigner's semicircle law, for a general class of potentials which we call "uniformly p-convex." To achieve this, we establish isoperimetric and transportation inequalities involving entropy with respect to the ensembles, which may be of independent interest.

[1] Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, England, United Kingdom; e-mail g.blower@lancaster.ac.uk

0022-4715/01/1000-0309$19.50/0

Let $M_n(\mathbb{R})$ denote the real $n \times n$ matrices, with symmetric part $M_n^s(\mathbb{R})$. Each $X \in M_n^s(\mathbb{R})$ has a unique list of eigenvalues $\lambda = (\lambda_j)_{j=1}^n$, in increasing order according to multiplicity, determining an element of the simplex

$$\mathbb{S}^n = \{(\lambda_j) \in \mathbb{R}^n : \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n\}. \tag{1.1}$$

The normalized trace on $M_n^s(\mathbb{R})$ is $\operatorname{trace}_n(X) = \frac{1}{n}\sum_{j=1}^n \lambda_j$, and the normalized Schatten $c^p(n)$ norms are defined on $M_n(\mathbb{R})$ by

$$\|X\|_{c^p(n)}^p = \operatorname{trace}_n(X^\dagger X)^{\frac{p}{2}} \qquad (X \in M_n(\mathbb{R})). \tag{1.2}$$

The normalized sequence space norms $\ell^p(n)$ on $\mathbb{R}^n$ are $\|\lambda\|_{\ell^p(n)} = (\frac{1}{n}\sum_{j=1}^n |\lambda_j|^p)^{1/p}$.

Let v be a continuous and even real function with $v(0) = 0$ and

$$v(x) \geqslant \frac{n+1}{n} \log |x| \tag{1.3}$$

for all sufficiently large $|x|$. Using the functional calculus of real symmetric matrices, we set $V(X) = \operatorname{trace}_n v(X)$. Taking dX to denote the product of Lebesgue measure on the entries lying on or above the leading diagonal of the symmetric matrices, we can define a bounded positive measure on $M_n^s(\mathbb{R})$ by $e^{-n^2 V(X)}\, dX$; then for some normalizing constant $0 < Z_n < \infty$,

$$\nu_n(dX) = Z_n^{-1} e^{-n^2 V(X)}\, dX \tag{1.4}$$

defines a probability measure on $M_n^s(\mathbb{R})$. This is termed the generalized orthogonal ensemble in ref. 7, and we now review its basic properties.

The group $O(n)$ of real orthogonal matrices acts on $M_n^s(\mathbb{R})$ by conjugation $X \mapsto UXU^\dagger$ $(X \in M_n^s(\mathbb{R}), U \in O(n))$, and the coset space may be identified with $\mathbb{S}^n$. This is merely a restatement of the fact that any real symmetric matrix may be diagonalized using real orthogonal matrices. Under this action, the measure ν_n is invariant in the sense that

$$\int_{M_n^s(\mathbb{R})} F(UXU^\dagger)\, \nu_n(dX) = \int_{M_n^s(\mathbb{R})} F(X)\, \nu_n(dX) \tag{1.5}$$

for all bounded and continuous functions $F\colon M_n^s(\mathbb{R}) \to \mathbb{R}$ and $U \in O(n)$. Thus the ensemble is statistically invariant under a change of basis of $\mathbb{R}^n$.

Let $\Lambda: X \mapsto (\lambda_j)$ be the map $M_n^s(\mathbb{R}) \to \mathbb{S}^n$ that associates to any matrix its ordered eigenvalues. The measure σ_n induced by Λ from ν_n is characterized by

$$\int_{\mathbb{S}^n} f(\lambda)\, \sigma_n(d\lambda) = \int_{M_n^s(\mathbb{R})} f(\Lambda(X))\, \nu_n(dX) \qquad (f \in C_b(\mathbb{S}^n)); \tag{1.6}$$

further, σ_n is absolutely continuous with respect to Lebesgue volume measure on $\mathbb{S}^n$ and may be represented as

$$\sigma_n(d\lambda) = Z(n, v)^{-1} \exp\left\{ \sum_{j,\,k:\, 1 \leqslant j < k \leqslant n} \log|\lambda_j - \lambda_k| - n \sum_{j=1}^{n} v(\lambda_j) \right\} d\lambda_1\, d\lambda_2 \cdots d\lambda_n \tag{1.7}$$

for some constant $0 < Z(n, v) < \infty$. One can show that there exists a probability measure $\mathbb{P}$ on the space $\Omega = \mathbb{R}^\infty$ of all matrix entries, and $X_n: (\Omega, \mathbb{P}) \to M_n^s(\mathbb{R})$ random matrices with distribution $\nu_n(dX)$ for $n \geqslant 1$.

The following physical interpretation is suggested by Wigner[20]: Let λ_j for $(j = 1, 2, \ldots, n)$ be the positions of atoms in a one-dimensional gas in a two-dimensional universe. Then $v(\lambda_j)$ represents the potential attracting the jth particle towards zero; whereas $\log|\lambda_j - \lambda_k|$ represents the electrostatic repulsion between particles j and k.

We make some special assumptions on the potential v which ensure that there is a single potential well given by a convex surface. When, as in Wigner's ensemble, we have $v(x) = x^2/2$, the corresponding V satisfies the parallelogram law $V(X) + V(Y) - 2V((X+Y)/2) = 2V((X-Y)/2)$; for our more general ensembles V satisfies a related inequality.

Definition (ref. 5). For $2 \leqslant p < \infty$ we shall say that V is uniformly p-convex if $V: M_n^s(\mathbb{R}) \to \mathbb{R}$ is convex and symmetric $V(X) = V(-X)$, and further there exist $C, \kappa_s > 0$, independent of dimension n, with the property

$$sV(X) + tV(Y) - V(sX + tY) \geqslant C(s + \kappa_s)\, \|X - Y\|^p_{c^p(n)} \qquad (X, Y \in M_n^s(\mathbb{R})) \tag{1.8}$$

for all $0 \leqslant s, t \leqslant 1$ with $s + t = 1$, where $\kappa_s/s \to 0$ as $s \to 0+$. Henceforth we shall assume that V has this property.

Let w be a real polynomial which defines an Orlicz function, so that:

(i) $w(0) = 0$, and w is strictly increasing and convex on $[0, \infty)$;

and further suppose

(ii) $xw'(x)/w(x)$ is increasing on $(0, \infty)$.

Proposition 1.1. The polynomial $v(x) = w(x^2)$ also satisfies (i) and (ii), and

$$V(X) = \mathrm{trace}_n\, w(X^\dagger X) \qquad (X \in M_n(\mathbb{R})) \tag{1.9}$$

is uniformly p-convex on $M_n(\mathbb{R})$, where p is the degree of v.

Examples. (a) The non-commutative Clarkson inequality[26] shows that, for $2 \leqslant p < \infty$, $v(x) = |x|^p$ gives rise to a uniformly p-convex potential $V(X) = \mathrm{trace}_n (X^\dagger X)^{\frac{p}{2}}$. In quantum field theory[7] one considers the class of even polynomials with non-negative coefficients, which satisfy (i) and (ii); and one can construct other examples. Brézin[8] *et al.* consider the potential $V(X) = \sum_{j=1}^n (\lambda_j^4 + b_2 \lambda_j^2/n)$ with $b_2 > 0$ in the context of the planar approximation to field theory, and show that the ground state of such an ensemble is related to the ground state of the one-dimensional x^4-anharmonic oscillator.

(b) When $v(x) = x^2$, we observe that $V(X) = \|X\|^2_{c^2(n)}$. Further, the scaled Hilbert–Schmidt norm reduces to the square sum of matrix entries $\|X\|^2_{c^2(n)} = \frac{1}{n} \sum_{j,k=1}^n [X]^2_{jk}$. Here (1.8) is a direct consequence of the parallelogram rule and ν_n is a Gaussian measure on $M_n^s(\mathbb{R})$. Indeed, the formula (1.4) factorizes as a tensor product of measures on the matrix entries, and one can regard $([X]_{jk})_{j \leqslant k}$ as mutually independent Gaussians with $[X]_{jj} \sim N(0, \frac{1}{2n})$ and $[X]_{jk} \sim N(0, \frac{1}{4n})$ for $j < k$.

For general uniformly p-convex V, we should not expect any simple description of the ensemble in terms of the matrix entries, which can be correlated as random variables. Mehta[20] argues that it is quite natural to consider ensembles which are invariant under orthogonal transformation, whereas it is somewhat artificial to assume that the matrix elements are statistically independent. The main purpose of this paper is to show that, nevertheless, several results known for ν_n and σ_n in the Gaussian case extend to the uniformly p-convex setting. See refs. 3, 4, 6, 15, 19 and 27 for corresponding Gaussian results.

The empirical distribution of eigenvalues is specified by the probability measures

$$\mu_n = \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j} \tag{1.10}$$

on $\mathbb{R}$, where δ_λ denotes the unit point mass at λ, and the λ_j are random eigenvalues subject to the distribution σ_n. One can also phrase results in terms of the normalized eigenvalue counting function for the ensemble σ_n, namely the random function $N_n(\lambda) = n^{-1}\#\{j: \lambda_j \leqslant \lambda\}$ ($\lambda \in \mathbb{R}$). Our extension of Wigner's semicircle law is:

Theorem 1.2. Let V be uniformly p-convex. Then under the laws (1.7) the empirical distributions of eigenvalues converge weakly almost surely to some non-random probability measure ρ supported in some bounded interval $[-K, K]$, so that

$$\int_{-\infty}^{\infty} f(x)\, \mu_n(dx) \to \int_{-(K+)}^{K+} f(x)\, \rho(dx) \qquad (n \to \infty) \tag{1.11}$$

almost surely for each bounded and continuous real function f, and further $N_n(\lambda) \to \int_{-\infty}^{\lambda+} \rho(dx)$ as $n \to \infty$ almost surely at all points of continuity of the limiting cumulative distribution function.

When v is twice continuously differentiable and convex, ρ is absolutely continuous and satisfies the singular integral equation

$$v'(\lambda) = \int_{-K}^{K} \frac{\rho(dx)}{\lambda - x} \qquad (-K < \lambda < K). \tag{1.12}$$

This may be inverted to give the density $d\rho/dx$ as a singular integral involving v'. See refs. 7, 22. In particular, Johansson[16] gives the following:

Proposition 1.3. Let v be as in Proposition 1.1. Then there exists a polynomial $r(x)$ of degree $p-2$, with zeros which are not real, such that

$$\rho(dx) = \frac{1}{\pi} r(x)\{(b-x)(x-a)\}^{1/2}\, \mathbb{I}_{[a,b]}(x)\, dx \tag{1.13}$$

where a and b are finite and $\mathbb{I}$ denotes the indicator function.

For Hölder continuous v satisfying $v(\lambda) \geqslant (2+\varepsilon) \log |\lambda|$ for $|\lambda| \to \infty$, Boutet de Monvel, Pastur and Shcherbina[7] analyze integrated densities of states, the probability measures (IDS) χ_n on $\mathbb{R}$ that satisfy $\int f\, d\chi_n = \iint f\, d\mu_n\, d\sigma_n$. They prove that $\int f\, d\chi_n \to \int f\, d\rho$ by mean field theory techniques, and also show that the $N_n(\lambda)$ converge in probability to $\int_{-\infty}^{\lambda+} d\rho$. Deift[9] *et al.* study a large class of C^∞ ensembles of complex Hermitian random matrices using Gaudin's method of orthogonal polynomials and achieve asymptotic results in the case in which the equilibrium measure is supported on a finite union of disjoint intervals. For continuous potentials with $v(\lambda) \geqslant (1+\varepsilon) \log |\lambda|$ ($|\lambda| \to \infty$) for some $\varepsilon > 0$, Kiessling and Spohn[18] establish the existence of a limiting eigenvalue distribution as the minimizer of a variational problem; this is unique under regularity hypotheses on v.

This variational approach, while very general, inevitably gives weaker control over the strength of convergence.

To obtain almost sure convergence, we use "concentration of measure," suggested by refs. 15, 27. Before stating the concentration inequalities, we recall some basic definitions from refs. 11, 12.

Let (Ω, d) be a complete and separable metric space, and ω and ϱ Radon probability measures on Ω. When ω is absolutely continuous with respect to ϱ, we can unambiguously define the relative entropy of ω with respect to ϱ by

$$\mathrm{Ent}(\omega \mid \varrho) = \int_\Omega \frac{d\omega}{d\varrho} \log \frac{d\omega}{d\varrho}\, d\varrho; \tag{1.14}$$

where by Jensen's inequality $0 \leqslant \mathrm{Ent}(\omega \mid \varrho) \leqslant \infty$.

When $\int_\Omega d(x, x_0)^p\, \varrho(dx) < \infty$ for some $x_0 \in \Omega$ and $p > 0$, we can define the transportation cost for cost function $d(x, y)^p$ by

$$\mathrm{W}_p(\omega, \varrho)^p = \inf\left\{ \iint_{\Omega\times\Omega} d(x, y)^p\, \pi(dx\, dy) : \pi \text{ has marginals } \pi_1 = \omega, \pi_2 = \varrho \right\}. \tag{1.15}$$

The infimum is attained, when Ω is locally compact; and the probability measure π defines a strategy for transporting the distribution ϱ to ω. The reader may wish to think of moving a pile of sand to fill a hole of equal volume as economically as possible when the cost of moving each grain grows like the pth power of the distance moved. For $p \geqslant 1$, $\mathrm{W}_p(\omega, \varrho)$ gives the Wasserstein metric on the space of probability distributions on Ω with finite pth moment. The Kantorovich–Rubinstein[11] duality formula gives the expression

$$\mathrm{W}_p(\omega, \varrho)^p = \sup\left\{ \int_\Omega f(x)\, \omega(dx) - \int_\Omega g(y)\, \varrho(dy) : \right.$$
$$\left. f(x) - g(y) \leqslant d(x, y)^p;\ x, y \in \Omega \right\} \tag{1.16}$$

wherein $f, g \colon \Omega \to \mathbb{R}$ may be taken to be continuous. For the empirical eigenvalue distribution we have the simplified identity from ref. 11

$$W_1(\mu_n, \rho) = \int_{-\infty}^{\infty} |N_n(\lambda) - \int_{-\infty}^{\lambda+} \rho(s)\, ds|\, d\lambda,$$

and we show this expression converges to zero almost surely as $n \to \infty$.

The following results are used in the proof of Theorem 1.2, and apply to a uniformly p-convex V.

Theorem 1.4. (a) Any probability measure ω on $M_n^s(\mathbb{R})$, which is absolutely continuous and of finite entropy with respect to v_n, satisfies the transportation inequality

$$W_p(\omega, v_n)^p \leqslant \frac{C_p}{n^2} \operatorname{Ent}(\omega \mid v_n) \tag{1.17}$$

for cost function $\|X-Y\|_{c^p(n)}^p$, where C_p depends upon p and the potential v only.

(b) Any probability measure ϱ on $\mathbb{S}^n$, which is absolutely continuous and of finite entropy with respect to σ_n, satisfies the transportation inequality corresponding to (1.17) for cost function $\|\lambda-\mu\|_{\ell^p(n)}^p = \frac{1}{n}\sum_{j=1}^n |\lambda_j-\mu_j|^p$.

We shall establish two closely related concentration inequalities for v_n and σ_n, the first of which is an isoperimetric inequality. For A a subset of a metric space (Ω, d), we define the ε-enlargement of A to be

$$A_\varepsilon = \{x \in \Omega : d(x, a) \leqslant \varepsilon \text{ for some } a \in A\}. \tag{1.18}$$

Theorem 1.5. (a) Let A be a Lebesgue measurable subset of $M_n^s(\mathbb{R})$ with $v_n(A) \geqslant \frac{1}{2}$. Then the ε-enlargement for the metric $\|X-Y\|_{c^p(n)}$ satisfies, for some constant $c_p > 0$,

$$v_n(A_\varepsilon) \geqslant 1-\exp(-c_p \varepsilon^p n^2) \qquad (\varepsilon > 0). \tag{1.19}$$

(b) Let $A \subseteq \mathbb{S}^n$ be a Lebesgue measurable set with $\sigma_n(A) \geqslant \frac{1}{2}$. Then its ε-enlargement for the metric of $\ell^p(n)$ satisfies, for some constant $c_p > 0$,

$$\sigma_n(A_\varepsilon) \geqslant 1-\exp(-c_p \varepsilon^p n^2) \qquad (\varepsilon > 0). \tag{1.20}$$

The functional form of the above isoperimetric inequality is a deviation estimate for Lipschitz functions. Given metric spaces (Ω_j, d_j) $(j=1,2)$, a function $f: \Omega_1 \to \Omega_2$ is said to be L-Lipschitz if $d_2(f(x), f(y)) \leqslant L d_1(x, y)$ for all $x, y \in \Omega_1$. We abusively call such an L the "Lipschitz norm" of f.

Theorem 1.6. (a) Let $F\colon (M_n^s(\mathbb{R}), \|\,.\,\|_{c^p(n)}) \to \mathbb{R}$ be a 1-Lipschitz function with $\int F(X)\, \nu_n(dX) = 0$. Then

$$\int_{M_n^s(\mathbb{R})} \exp\{tF(X)\}\, \nu_n(dX) \leqslant \exp\{t^q c_p n^{-2/(p-1)}\} \qquad (t > 0) \tag{1.21}$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and c_p depends upon p and v only.

(b) Any 1-Lipschitz function $f\colon (\mathbb{S}^n, \|\,.\,\|_{\ell^p(n)}) \to \mathbb{R}$ with $\int f(\lambda)\, \sigma_n(d\lambda) = 0$ satisfies such an inequality, with similar constants.

As p increases, so the concentration inequality becomes stronger.

The proof of these results begins in section two, where we derive some basic properties of uniformly convex potentials. In section three we shall show that Theorems 1.4(a), 1.5(a) and 1.6(a) may be deduced from the Prékopa–Leindler inequality by the procedure of Bobkov and Ledoux[5]. Theorems 1.4(b), 1.5(b) and 1.6(b) will be deduced in section four by consideration of the eigenvalue map $\Lambda\colon X \mapsto (\lambda_j)$. In section five we use the concentration inequalities to establish almost sure convergence in the weak sense for the empirical distribution of eigenvalues. Section six contains another approach to Theorem 1.2 for 2-uniformly convex potentials, which emphasizes the close connection between Theorems 1.4 and 1.6 and known results from the theory of logarithmic Sobolev inequalities as developed in refs. 1, 13, 23. In section seven we consider perturbations of a uniformly p-convex potential and how the limiting measure changes under perturbation. We conclude the paper with a conjectured extension of Theorem 1.2 to multiple-well potentials.

2. UNIFORMLY CONVEX POTENTIALS

Proof of Proposition 1.1. Clarkson showed that for $2 \leqslant p < \infty$, functions f and g in $L^p(\mathbb{R})$ satisfy $\int (|f|^p + |g|^p) - 2\int |(f+g)/2|^p \geqslant 2\int |(f-g)/2|^p$; this expresses the uniform p-convexity of $L^p(\mathbb{R})$. Geometrically it means that the unit sphere $\{f \in L^p(\mathbb{R}) : \int |f|^p = 1\}$ has no flat faces. Similar inequalities hold for the Schatten norms of matrices $\|X\|^p_{c^p(n)} = \mathrm{trace}_n(X^\dagger X)^{p/2}$, as was demonstrated by Dixmier *et al.*[28] In ref. 2, Ball, Carlen and Lieb give a historical account of such non-commutative Clarkson inequalities, and present best possible constants for the full range of p.

In this section we shall give a self-contained proof that

$$V(X) + V(Y) - 2V\left(\frac{X+Y}{2}\right) \geqslant 2V\left(\frac{X-Y}{2}\right). \tag{2.1}$$

This implies uniform p-convexity, as we now verify. The inequality

$$sV(X)+tV(Y)-V(sX+tY)\geqslant \min\{t,s\}\,V\left(\frac{X-Y}{2}\right) \tag{2.2}$$

then holds for all $0\leqslant s,\ t\leqslant 1$ with $s+t=1$. Indeed, it holds for $s=0$, and $s=\frac{1}{2}$ by (2.1); the expression $V(tX+sY)$ is a convex function of s, and the right-hand side of (2.2) is linear in s for $0\leqslant s\leqslant \frac{1}{2}$, so the inequality persists for $0\leqslant s\leqslant \frac{1}{2}$ by concavity of the left-hand side.

It is a simple exercise to show that $v(x)$ satisfies (i) and (ii) of Proposition 1.1, if and only if w does. To bound below the right-hand side of (2.2), we observe that $v(x)/v(tx)$ is a decreasing function of $x\in(0,\infty)$ for each fixed $0<t<1$; so by considering the limiting behaviour at ∞, we deduce that $v(tx)\geqslant t^p v(x)$. In particular, we have $v(x/2)\geqslant 2^{-p}v(x)$, which implies $V(\frac{X-Y}{2})\geqslant 2^{-p}V(X-Y)$. In the language of Orlicz space theory, we have shown that v satisfies the Δ_2 condition at zero.

The expression $v(x)/x^p$ is decreasing on account of (ii), and hence bounded below by its limit at ∞ which is the leading coefficient a_p of v. Hence

$$V\left(\frac{X-Y}{2}\right)\geqslant \frac{a_p}{2^p}\,\|X-Y\|^p_{c^p(n)}. \tag{2.3}$$

To justify (2.1), we follow the proof of the non-commutative Clarkson inequalities from ref. 26. For any matrix X, we let μ_j be the jth largest singular number of X; that is, the jth largest eigenvalue of $(X^\dagger X)^{1/2}$ counted according to multiplicity. With $W(X)=\frac{1}{n}\sum_{j=1}^{n} w(\mu_j(X))$ we have, for any matrices A and B,

$$W\left(\frac{A+B}{2}\right)\leqslant \frac{1}{2}\,(W(A)+W(B)). \tag{2.4}$$

Indeed, by Lidskii's Theorem,[26] there exists a doubly substochastic matrix $[a_{km}]$ such that

$$\mu_k\left(\frac{1}{2}\,(A+B)\right)=\frac{1}{2}\,\mu_k(A)+\frac{1}{2}\sum_{m=1}^{n} a_{km}\mu_m(B)\qquad (k\geqslant 1). \tag{2.5}$$

Since w is convex and increasing, we deduce

$$w\left\{\mu_k\left(\frac{1}{2}\,(A+B)\right)\right\}\leqslant \frac{1}{2}\,w(\mu_k(A))+\frac{1}{2}\sum_{m=1}^{n} |a_{km}|\,w(\mu_m(B))\qquad (k\geqslant 1) \tag{2.6}$$

and we now deduce (2.4) on summing (2.6) over k.

Conversely, for positive matrices A and B we have

$$W(A)+W(B) \leqslant W(A+B), \tag{2.7}$$

as we now recall. By adding εI_n to A if necessary, we can assume that $A+B$ is invertible. The matrices

$$C = A^{\frac{1}{2}}(A+B)^{-\frac{1}{2}} \quad \text{and} \quad D = B^{\frac{1}{2}}(A+B)^{-\frac{1}{2}} \tag{2.8}$$

satisfy $C^*C+D^*D = I_n$, which already implies $\|C^*\|, \|D^*\| \leqslant 1$, and furthermore

$$A = C(A+B)\, C^* \quad \text{and} \quad B = D(A+B)\, D^*. \tag{2.9}$$

Let (ϕ_k) be an orthonormal basis consisting of eigenvectors of A corresponding to eigenvalues $(\mu_k(A))$, and (ψ_k) an orthonormal basis consisting of eigenvectors of B corresponding to eigenvalues $(\mu_k(B))$. It follows from Jensen's inequality and the spectral theorem that

$$\begin{aligned} w(\mu_k(A)) &= w(\langle A\phi_k, \phi_k\rangle) \leqslant \langle w(A+B)\, C^*\phi_k, C^*\phi_k\rangle, \\ w(\mu_k(B)) &= w(\langle B\psi_k, \psi_k\rangle) \leqslant \langle w(A+B)\, D^*\psi_k, D^*\psi_k\rangle. \end{aligned} \tag{2.10}$$

On summing these expressions over k, we deduce that

$$\begin{aligned} W(A)+W(B) &\leqslant \mathrm{trace}_n(Cw(A+B)\, C^* + Dw(A+B)\, D^*) \\ &= \mathrm{trace}_n((C^*C+D^*D)\, w(A+B)) \\ &= \mathrm{trace}_n(w(A+B)). \end{aligned} \tag{2.11}$$

Since $A+B \geqslant 0$, its singular numbers are simply its eigenvalues; and so the spectral theorem gives us (2.7).

To complete the proof of (2.1), we recall that $V(X) = \mathrm{trace}_n\, w(X^\dagger X)$, and use the inequality (2.4) to deduce

$$\begin{aligned} V(X)+V(Y) = W(X^\dagger X)+W(Y^\dagger Y) &\geqslant 2W\left(\frac{X^\dagger X+Y^\dagger Y}{2}\right) \\ &= 2W\left(\frac{1}{4}(X+Y)^\dagger\,(X+Y)+\frac{1}{4}(X-Y)^\dagger\,(X-Y)\right). \end{aligned} \tag{2.12}$$

On account of the inequality (2.7), this is

$$\geqslant 2W\left(\frac{1}{4}(X+Y)^{\dagger}(X+Y)\right)+2W\left(\frac{1}{4}(X-Y)^{\dagger}(X-Y)\right)$$

$$=2V\left(\frac{X+Y}{2}\right)+2V\left(\frac{X-Y}{2}\right). \tag{2.13}$$

3. CONCENTRATION OF MEASURE FOR MATRIX ENSEMBLE

The space $M_n^s(\mathbb{R})$ may be identified with $\mathbb{R}^{n(n+1)/2}$ by associating to each matrix X the vector $([X]_{jk})$ $(j\leqslant k)$ of its entries on or above the leading diagonal. The Prékopa–Leindler inequality asserts that, if u, v, w: $\mathbb{R}^{n(n+1)/2}\to\mathbb{R}_+$ are measurable functions with $w(tX+sY)\geqslant u(X)^t\,v(Y)^s$ for all $X, Y\in\mathbb{R}^{n(n+1)/2}$ and $s, t\geqslant 0$ with $s+t=1$, then

$$\int w\,dX\geqslant\left(\int u\,dX\right)^t\left(\int v\,dX\right)^s \tag{3.1}$$

where dX denotes Lebesgue measure on $\mathbb{R}^{n(n+1)/2}$. From this fundamental inequality all of the results of this section follow by the procedure which Bobkov and Ledoux[5] have developed.

Proof of Theorem 1.4(a). Employing a device of Maurey[5], we let

$$L_s(X,Y)=\frac{1}{ts}\left(tV(X)+sV(Y)-V(tX+sY)\right)\qquad (X, Y\in M_n^s(\mathbb{R})) \tag{3.2}$$

for $s, t\geqslant 0$ with $s+t=1$. It follows from the uniform p-convexity of V that

$$L_s(X,Y)\geqslant\frac{C}{pt}(1+\kappa_s)\,\|X-Y\|^p_{c^p(n)}\qquad (X, Y\in M_n^s(\mathbb{R})) \tag{3.3}$$

where $\kappa_s\to 0$ as $s\to 0+$ and all the positive constants are independent of n, X and Y. Let F, G: $M_n^s(\mathbb{R})\to\mathbb{R}$ be bounded and continuous functions with

$$F(Y)-G(X)\leqslant\frac{C}{p}\|X-Y\|^p_{c^p(n)}\qquad (X, Y\in M_n^s(\mathbb{R})). \tag{3.4}$$

Then by the Prékopa–Leindler inequality

$$1=\int e^{-n^2V}dX/Z_n \geqslant \left(\int e^{-sn^2G-n^2V}dX/Z_n\right)^{1/s}\left(\int e^{tn^2F-n^2V}dX/Z_n\right)^{1/t}. \tag{3.5}$$

Letting $s\to 0+$, we deduce that

$$\int \exp(n^2F)\,d\nu_n \leqslant \exp\int n^2G\,d\nu_n. \tag{3.6}$$

The relative entropy of an absolutely continuous probability measure ϱ with respect to ν_n may be expressed as

$$\mathrm{Ent}(\varrho\mid\nu_n)=\sup\left\{\int h\,d\varrho : \int e^h\,d\nu_n\leqslant 1\right\}. \tag{3.7}$$

This duality relation is a consequence of the elementary fact that $x\log x$ $(x>0)$ attains its minimum at $x=1/e$. By the preceding inequality (3.6), $h(X)=n^2F(X)-n^2\int G\,d\nu_n$ is an admissible choice to substitute into the right-hand side of (3.7); and so we deduce

$$\mathrm{Ent}(\varrho\mid\nu_n)\geqslant n^2\int F(X)\,\varrho(dX)-n^2\int G(Y)\,\nu_n(dY). \tag{3.8}$$

We conclude the argument by taking the supremum of (3.8) over all pairs (F,G) subject to the constraint (3.4). The Kantorovich–Rubinstein duality theorem of (1.16) shows that

$$\mathrm{Ent}(\varrho\mid\nu_n)\geqslant\frac{n^2p}{C}\,\mathrm{W}_p(\varrho,\nu_n)^p. \tag{3.9}$$

Proof of Theorem 1.5(a). Marton[27] showed that transportation inequalities may be converted into isoperimetric inequalities by the following argument. Let $B=(A_\varepsilon)^c$ be the complement of the ε-enlargement of A in $M_n^s(\mathbb{R})$, and introduce the conditional probabilities

$$\varrho_A(E)=\nu_n(A\cap E)/\nu_n(A),\qquad \varrho_B(E)=\nu_n(B\cap E)/\nu_n(B), \tag{3.10}$$

which are supported on the sets A and B at distance ε apart. It follows that

$$\varepsilon\leqslant \mathrm{W}_p(\varrho_A,\varrho_B)\leqslant \mathrm{W}_p(\varrho_A,\nu_n)+\mathrm{W}_p(\nu_n,\varrho_B) \tag{3.11}$$

by the triangle inequality; and the transportation inequality (3.9) shows this sum to be

$$\leqslant \left(\frac{C}{pn^2}\operatorname{Ent}(\varrho_A \mid \nu_n)\right)^{1/p} + \left(\frac{C}{pn^2}\operatorname{Ent}(\varrho_B \mid \nu_n)\right)^{1/p}; \tag{3.12}$$

and by calculation we see this expression is

$$\leqslant \left(\frac{C}{pn^2}\log\frac{1}{\nu_n(A)}\right)^{1/p} + \left(\frac{C}{pn^2}\log\frac{1}{\nu_n(B)}\right)^{1/p}. \tag{3.13}$$

From this inequality (3.13), and the obvious fact $\nu_n(B) \leqslant 1-\nu_n(A)$, we deduce

$$\nu_n(A)\,\nu_n(B) \leqslant \min\left\{\frac{1}{4}, \exp\left(-\frac{\varepsilon^p n^2 p}{2^{p-1}C}\right)\right\} \leqslant \frac{1}{2}\exp\left(-\frac{\varepsilon^p pn^2}{2^p C}\right); \tag{3.14}$$

which implies the required result.

Proof of Theorem 1.6(a). The condition, that inequality

$$\int_{M_n^s(\mathbb{R})} e^{tG-t^q\kappa}\nu_n(dX) \leqslant 1 \qquad (t>0) \tag{3.15}$$

holds for all 1-Lipschitz $G\colon M_n^s(\mathbb{R}) \to \mathbb{R}$ with $\int G(X)\,\nu_n(dX) = 0$, is equivalent to the condition that inequality

$$\operatorname{Ent}(\varrho \mid \nu_n) \geqslant t\int_{M_n^s(\mathbb{R})} G\,d\varrho - t^q\kappa \qquad (t>0) \tag{3.16}$$

holds for all such G and all probabilities ϱ of finite entropy with respect to ν_n. We recall the proof from ref. 4.

The forward implication is immediate from the duality formula (3.7); whereas the reverse implication is given by the following proof by contradiction: Suppose that (3.16) holds and $\int e^{tG-t^q\kappa}\,d\nu_n = e^{\alpha}$, where $\alpha > 0$. Then $d\varrho = e^{tG-t^q\kappa-\alpha}\,d\nu_n$ defines a probability for which the entropy is

$$\operatorname{Ent}(\varrho \mid \nu_n) = \int_{M_n^s(\mathbb{R})} (tG - t^q\kappa - \alpha)\,d\varrho \leqslant \operatorname{Ent}(\varrho \mid \nu_n) - \alpha \tag{3.17}$$

by (3.16). Since $\alpha > 0$, this cannot hold.

We further reduce the claim of the Theorem by optimizing (3.16) over t, which leaves us with the equivalent condition

$$\mathrm{Ent}(\varrho \mid \nu_n) \geqslant \kappa^{-\frac{1}{q-1}} \left(\int_{M_n^s(\mathbb{R})} G d\varrho \right)^p q^{-\frac{p}{q}} p^{-1}. \tag{3.18}$$

However, the transportation inequality (1.17) implies by Hölder's inequality that

$$\mathrm{Ent}(\varrho \mid \nu_n) \geqslant \frac{n^2 p}{C} \mathrm{W}_1(\varrho, \nu_n)^p, \tag{3.19}$$

where the cost function is now $\|X-Y\|_{c^p(n)}$. By duality (1.16) we have

$$\mathrm{Ent}(\varrho \mid \nu_n) \geqslant \frac{n^2 p}{C} \left(\int_{M_n^s(\mathbb{R})} G \, d\varrho \right)^p \tag{3.20}$$

for any G as above. We deduce that the concentration inequality (3.15) holds with constant

$$\kappa = \frac{1}{q} \left(\frac{\sqrt{C}}{pn} \right)^{\frac{2}{p-1}}. \tag{3.21}$$

4. CONCENTRATION OF EIGENVALUE DISTRIBUTION

Lemma 4.1. The map Λ, which associates to each real symmetric matrix its ordered list of eigenvalues, is 1-Lipschitz for the norms $c^p(n) \to \ell^p(n)$.

Proof. Lidskii[26] showed that, for real symmetric matrices X and Y, there exists a doubly substochastic matrix $[a_{jk}]$ for which the respective eigenvalues satisfy

$$\lambda_j(X) - \lambda_j(Y) = \sum_{k=1}^{n} a_{jk} \lambda_k(X-Y) \qquad (j = 1, 2, \ldots, n). \tag{4.1}$$

Every doubly stochastic matrix is an absolute convex combination of permutation matrices, and hence by convexity we have the required inequality

$$\left(\frac{1}{n} \sum_{j=1}^{n} |\lambda_j(X) - \lambda_j(Y)|^p \right)^{1/p} \leqslant \left(\frac{1}{n} \sum_{k=1}^{n} |\lambda_k(X-Y)|^p \right)^{1/p}. \tag{4.2}$$

Proof of Theorem 1.4(b). Using this lemma, we can deduce Theorem 1.4(b) from Theorem 1.4(a) by abstract duality arguments from refs. 11, 14. Let $m_{O(n)}$ be Haar probability measure on the real orthogonal $n \times n$ matrices, and let σ_n and ϱ be measures on $\mathbb{S}^n$, as in the theorem. Abusing notation, we write $\Lambda(X)$ for the diagonal matrix for which the diagonal entries are the ordered eigenvalues of $X \in M_n^s(\mathbb{R})$. We denote by $\tilde{\varrho}$ the probability measure on $M_n^s(\mathbb{R})$ that satisfies

$$\int_{M_n^s(\mathbb{R})} F(X)\, \tilde{\varrho}(dX) = \iint_{O(n) \times \mathbb{S}^n} F(U\, \Lambda(X)\, U^\dagger)\, m_{O(n)}(dU)\, \varrho(d\lambda) \tag{4.3}$$

for all continuous and bounded functions $F: M_n^s(\mathbb{R}) \to \mathbb{R}$. Since $U\, \Lambda(X)\, U^\dagger$ has the same eigenvalues as $\Lambda(X)$ it is easy to check that ϱ is the measure on $\mathbb{S}^n$ induced from $\tilde{\varrho}$ by the map $\Lambda: M_n^s(\mathbb{R}) \to \mathbb{S}^n$; and we recall that σ_n is induced by ν_n.

By the Kantorovich–Rubinstein duality formula (1.16), we have

$$\mathrm{W}_p(\varrho, \sigma_n)^p = \sup \left\{ \int_{\mathbb{S}^n} f(\lambda)\, \varrho(d\lambda) - \int_{\mathbb{S}^n} g(\lambda')\, \sigma_n(d\lambda') : \right.$$
$$\left. f(\lambda) - g(\lambda') \leqslant \|\lambda - \lambda'\|^p_{\ell^p(n)} \right\};$$

and using the definition of the induced measure together with Lemma 4.1, we see this is

$$\leqslant \sup \left\{ \int_{M_n^s(\mathbb{R})} f(\Lambda(X))\, \tilde{\varrho}(dX) - \int_{M_n^s(\mathbb{R})} g(\Lambda(Y))\, \nu_n(dY) : \right.$$
$$\left. f(\Lambda(X)) - g(\Lambda(Y)) \leqslant \|X - Y\|^p_{c^p(n)} \right\}. \tag{4.4}$$

Applying the duality principle again, we deduce this is

$$\leqslant \mathrm{W}_p(\tilde{\varrho}, \nu_n)^p. \tag{4.5}$$

By Theorem 1.4(a), this transportation cost is at most $c_p n^{-2}\, \mathrm{Ent}(\tilde{\varrho} \mid \nu_n)$; and so to conclude the proof of Theorem 1.4(b), it suffices to show $\mathrm{Ent}(\tilde{\varrho} \mid \nu_n) = \mathrm{Ent}(\varrho \mid \sigma_n)$. This follows from the fact that the measures $\tilde{\varrho}$ and ν_n are invariant under the dual of the conjugation action of $O(n)$ which rotates the function $h: M_n(\mathbb{R}) \to \mathbb{R}$ to $h^U(X) = h(UXU^\dagger)$ $(X \in M_n^s(\mathbb{R}),$ $U \in O(n))$. Indeed, there exists a continuous h which satisfies

$$\int_{M_n^s(\mathbb{R})} e^{h(X)} \nu_n(dX) \leqslant 1 \tag{4.6}$$

and

$$\int_{M_n^s(\mathbb{R})} h(X)\,\tilde{\varrho}(dX) \geqslant \mathrm{Ent}(\tilde{\varrho}\mid\nu_n) - \varepsilon; \tag{4.7}$$

moreover, by invariance the same conditions hold for h^U. Averaging (4.6) over $U \in O(n)$ with respect to Haar measure, we see that the invariant function $\langle h\rangle(X) = \int_{O(n)} h^U(X)\, m_{O(n)}(dU)$ satisfies

$$\int_{M_n^s(\mathbb{R})} \langle h\rangle(X)\,\tilde{\varrho}(dX) \geqslant \mathrm{Ent}(\tilde{\varrho}\mid\nu_n) - \varepsilon \tag{4.8}$$

by Fubini's Theorem; whereas by Jensen's inequality $\int e^{\langle h\rangle}\nu_n(dX) \leqslant 1$. Since $\langle h\rangle$ may be regarded as a function on $\mathbb{S}^n$, and $\tilde{\varrho}$ and ν_n induce ϱ and σ_n respectively, we deduce

$$\mathrm{Ent}(\varrho\mid\sigma_n) \geqslant \mathrm{Ent}(\tilde{\varrho}\mid\nu_n).$$

The converse inequality is clear.

Proofs of Theorems 1.5(b) and 1.6(b). Using the arguments of the preceding section, we can deduce Theorems 1.5(b) and 1.6(b) from 1.4(b).

We shall now show that the measure ν_n is concentrated near to zero. Ultimately our estimates will show that the weak limit of the empirical distribution is of compact support.

Proposition 4.2. There exist constants $0 < \alpha_p, K_p < \infty$, independent of n, such that

$$\int_{M_n^s(\mathbb{R})} \exp\{\alpha_p\, \|X\|^p_{c^p(n)}\}\,\nu_n(dX) \leqslant K_p. \tag{4.9}$$

Proof. By a standard convexity lemma from refs. 24, 21 p.7, there exists a net of the unit sphere in $c^q(n)$ consisting of $n\times n$ matrices $Y_1, Y_2, \ldots, Y_N$, where $N \leqslant 5^{n^2}$, such that $\|Y_j\|_{c^q(n)} = 1$ and

$$\|X\|_{c^p(n)} \leqslant 2\max_j \mathrm{trace}_n(XY_j) \qquad (X \in M_n^s(\mathbb{R})). \tag{4.10}$$

The functional $F_j\colon X \mapsto \mathrm{trace}_n(XY_j)$ has Lipschitz norm one $c^p(n) \to \mathbb{R}$; and $\int F_j(X)\,\nu_n(dX) = 0$, since ν_n is symmetric. By Theorem 1.6(a) we have

$$\int_{M_n^s(\mathbb{R})} \exp\{tF_j(X)\}\,\nu_n(dX) \leqslant \exp\{t^q n^{-2/(p-1)} c_p\} \qquad (t > 0), \tag{4.11}$$

from which we deduce by Chebyshev's inequality

$$\nu_n[F_j(X) > s] \leqslant \exp\left\{ -\frac{s^p n^2 (q-1)}{c_p^{p-1} q^p} \right\} \qquad (s > 0). \tag{4.12}$$

We have, by the choice of functionals in (4.10),

$$\begin{aligned} \nu_n[\|X\|_{c^p(n)} > 2s] &\leqslant \sum_{j=1}^{N} \nu_n[F_j(X) > s] \\ &\leqslant \exp\left\{ n^2 \log 5 - \frac{s^p n^2 (q-1)}{c_p^{p-1} q^p} \right\}. \end{aligned} \tag{4.13}$$

We choose s_0 so that $\log 5 = s_0^p (q-1)/q^q c_p^{p-1}$, and then with $s_1 = 2s_0$ we have $\beta_p > 0$ with

$$\frac{1}{n^2} \log \nu_n[\|X\|_{c^p(n)} > s] \leqslant -\beta_p (s - s_1)^p \qquad (s \geqslant s_1). \tag{4.14}$$

The stated result now follows by a straightforward application of Fubini's Theorem. For a general discussion of deviation inequalities of the style (4.14), see Ellis' book.[12]

Chevet[24, 29] proved that if Y is a $n \times n$ matrix whose entries are mutually independent Gaussian $N(0, 1/n)$ random variables, then there exists a constant K independent of n for which the mean operator norm satisfies $\mathbb{E}\,\|Y\| \leqslant K$. The proof of this result uses a fact, special to Gaussian processes, known as Slepian's Lemma:

If $(Z_j: 1 \leqslant j \leqslant N)$ and $(Z'_j: 1 \leqslant j \leqslant N)$ are Gaussian processes with $\mathbb{E}\,|Z_j - Z_k|^2 \leqslant \mathbb{E}\,|Z'_j - Z'_k|^2$ for all j and k, then $\mathbb{E}\sup_j Z_j \leqslant 2\mathbb{E}\sup_j Z'_j$.

For general orthogonal ensembles there is no such simple connection between covariance and bounds on the associated processes; in Section 6 we investigate how uniform 2-convexity is related to derivatives of the potential and to statistical properties of the eigenvalues. Here we use concentration of measure to prove an analogue of Chevet's Lemma in the uniformly p-convex case.

Proposition 4.3. The means of the largest eigenvalue are uniformly bounded: there exists $K < \infty$, independent of n, such that

$$\int_{\mathbb{S}^n} \lambda_n \sigma_n(d\lambda) \leqslant K. \tag{4.15}$$

Proof. The numerical radius of a real symmetric matrix co-incides with its operator norm, which is the largest eigenvalue in modulus. We can choose a net $(\xi_j)_{j=1}^n$ of the unit sphere of $\ell^2(n)$, of cardinality at most $N=5^n$, such that $\lambda_n(X)\leqslant 2\max_j F_j(X)$, where

$$F_j(X)=\langle X\xi_j,\xi_j\rangle_{\ell^2(n)} \qquad (X\in M_n^s(\mathbb{R});\, j=1,2,\ldots,N) \tag{4.16}$$

are the associated linear functionals. We note that $\int F_j(X)\,\nu_n(dX)=0$ by symmetry, and the Lipschitz norm of F_j is at most $n^{1/p}$ for $j=1,2,\ldots,N$.

By the concentration inequality of Theorem 1.6(a), we have

$$\int_{M_n^s(\mathbb{R})}\exp\{sF_j(X)\}\,\nu_n(dX)\leqslant\exp\left\{c_p s^q n^{-\frac{1}{p-1}}\right\} \qquad (s>0); \tag{4.17}$$

and consequently by Chebyshev's inequality

$$\nu_n[F_j>t]\leqslant\exp\left\{-\frac{(q-1)\,nt^p}{q^p c_p^{p-1}}\right\} \qquad (t>0). \tag{4.18}$$

From the choice of F_j, we deduce that

$$\sigma_n[\lambda_n>2t]\leqslant\sum_{j=1}^N \nu_n[F_j>t]\leqslant 5^n\exp\left\{-\frac{(q-1)\,nt^p}{q^p c_p^{p-1}}\right\}. \tag{4.19}$$

From this deviation estimate the stated results follows easily by Fubini's Theorem.

5. CONVERGENCE OF THE EMPIRICAL DISTRIBUTION

Let $\mu_n=\frac{1}{n}\sum_{j=1}^n\delta_{\lambda_j}$ be the empirical distribution of states, where the eigenvalues λ_j have joint distribution σ_n. We shall show that the μ_n are weak-$*$ convergent almost surely as $n\to\infty$ to a compactly supported non-random distribution. The structure of the argument comes from Haagerup and Thorbjørnsen's paper,[15] and they attribute it to Pisier. The results of ref. 15 apply to the Gaussian case which involves the uniformly 2 convex Hilbert–Schmidt norm on matrices. We extend the method to uniformly p-convex potentials.

To show that $\int f\,d\mu_n$ converges for $f\in C_0(\mathbb{R})$, we take f belonging to the linear subspace of Lipschitz functions, which is dense for the supremum norm, and consider $F_n(\lambda)=\frac{1}{n}\sum_{j=1}^n f(\lambda_j)$. Throughout this section, $\mathbb{S}^n$ shall have the $\ell^p(n)$ metric and $M_n^s(\mathbb{R})$ the $c^p(n)$ metric. It follows from Hölder's inequality that the Lipschitz norm of F is at most L, if the Lipschitz norm of f is at most L.

Regarding F_n as a random variable, we let m_n be a median of F_n, so that $A = [F_n \leqslant m_n]$ has $\lim_{x \to m_n -} \sigma_n[F_n \leqslant x] \leqslant \frac{1}{2} \leqslant \sigma_n(A)$. Since F_n is L-Lipschitz, we have

$$[-\varepsilon + m_n \leqslant F_n \leqslant m_n + \varepsilon] \supseteq A_{\varepsilon/L} \cap (A^c)_{\varepsilon/L} \tag{5.1}$$

and hence, by the isoperimetric inequality of Theorem 1.5(b),

$$\sigma_n[-\varepsilon + m_n \leqslant F_n \leqslant m_n + \varepsilon] \geqslant 1 - 2\exp\{-c_p n^2 \varepsilon^p / L^p\} \qquad (\varepsilon > 0). \tag{5.2}$$

By the first Borel-Cantelli Lemma we deduce that

$$\mathbb{P}[|F_n - m_n| > \varepsilon \text{ for infinitely many } n] = 0 \tag{5.3}$$

since the series $\sum_{n=1}^{\infty} \exp\{-c_p n^2 \varepsilon^p / L^p\}$ is rapidly convergent. This shows that $F_n - m_n \to 0$ almost surely as $n \to \infty$.

By a further application of the isoperimetric inequality, we can replace the medians by the means $M_n = \iint f \, d\mu_n \, d\sigma_n$. Indeed, for $0 < \alpha < \infty$ the sets $B_n = [-\varepsilon/n^\alpha + m_n \leqslant F_n \leqslant m_n + \varepsilon/n^\alpha]$ have $\sigma_n(B_n) \to 1$ as $n \to \infty$, as in (5.2). By Fubini's theorem we have

$$|M_n - m_n| \leqslant \int_{\mathbb{S}^n} |F(\lambda) - m_n| \, \sigma_n(d\lambda) \tag{5.4}$$

$$\leqslant \int_{B_n} |F(\lambda) - m_n| \, \sigma_n(d\lambda) + \int_0^\infty \sigma_n \left[|F_n - m_n| \geqslant r + \frac{\varepsilon}{n^\alpha} \right] dr. \tag{5.5}$$

The first integral in (5.5) is plainly at most $L\varepsilon/n^\alpha$, and the second integral involves

$$\sigma_n \left[|F_n - m_n| \geqslant r + \frac{\varepsilon}{n^\alpha} \right] \leqslant 1 - \sigma_n[(B_n)_{r/L}] \tag{5.6}$$

where both these estimates follow from the L-Lipschitz bound on F_n. It follows from the isoperimetric inequality (1.20) that the final integral in (5.5) is at most

$$\begin{aligned} \int_0^\infty \left(1 - \sigma_n[(B_n)_{r/L}]\right) dr &\leqslant \int_0^\infty \exp\left\{-c_p n^2 r^p / L^p\right\} dr \\ &= \frac{L}{p(c_p n^2)^{\frac{1}{p}}} \Gamma\left(\frac{1}{p}\right) \end{aligned} \tag{5.7}$$

for all sufficiently large n.

By Theorem 2 of ref. 7, the M_n converge as $n \to \infty$ since the integrated densities of states converge weakly. We deduce that $f \mapsto \lim_n \int f\, d\mu_n$ defines a positive linear functional on $C[-K', K']$ of norm at most one, for each $K' < \infty$. Since μ_n is supported on $[\lambda_1, \lambda_n]$, where $\lambda_1 \sim -\lambda_n$ and λ_n obeys the distributional estimate (4.19), we deduce from the F. Riesz representation theorem that there exists a probability measure ρ supported on $[-K, K]$ for some $K < \infty$ for which (1.11) holds.

6. UNIFORM 2-CONVEXITY

In this section we give an alternative approach to the inequalities which are used in the proof of Theorem 1.2, using hypotheses familiar from the theory of logarithmic Sobolev inequalities as in refs. 1, 13. Let v be an even and three-times continuously differentiable real function and take $V(X) = \mathrm{trace}_n\, v(X)$ for each $X \in M_n^s(\mathbb{R})$. We can regard V as a function of the matrix entries $[X]_{jk}$ $(1 \leqslant j \leqslant k \leqslant n)$, and form the gradient ∇V, represented by the matrix of first-order partial derivatives $\partial_{jk} V$ with respect to the $[X]_{jk}$, with $\partial_{jk} V = \partial_{kj} V$; this pairs naturally with $Y \in M_n^s(\mathbb{R})$ according to $\langle \nabla V \mid Y \rangle = \sum_{1 \leqslant j,k \leqslant n} (\partial_{jk} V)[Y]_{jk}$. The Hessian of V is the doubly indexed matrix $\partial_{jk}\partial_{\ell m} V$ of second -order partial derivatives with respect to $[X]_{jk}$, and this may be associated with a bilinear form $(Y, W) \mapsto \sum_{1 \leqslant j,k,\ell,m \leqslant n} (\partial_{jk}\partial_{\ell m} V)[Y]_{jk}\,[W]_{\ell m}$ on $M_n^s(\mathbb{R}) \times M_n^s(\mathbb{R})$. In this section we prove:

Theorem 6.1. Suppose that there exists $\delta > 0$, with $v''(x) \geqslant \delta > 0$ for all $x \in \mathbb{R}$. Then the empirical distribution of eigenvalues, associated with the ensemble on $M_n^s(\mathbb{R})$ given by

$$\nu_n(dX) = Z_n^{-1} \exp(-n^2 V(X))\, dX, \tag{6.1}$$

converges weakly almost surely as $n \to \infty$ to some non-random probability measure supported in a bounded interval in $\mathbb{R}$.

The proof depends upon the following two lemmas.

Lemma 6.2. The function $V \colon M_n^s(\mathbb{R}) \to \mathbb{R}$ is twice continuously differentiable with

$$\langle \mathrm{Hess}\, V, Y \otimes Y \rangle \geqslant \delta\, \|Y\|_{c^2(n)}^2 \qquad (Y \in M_n^s(\mathbb{R})). \tag{6.2}$$

Proof of Lemma 6.2. The spectra of the operators $X + tY$ with X and Y in $M_n^s(\mathbb{R})$ and $-1 < t < 1$ lie in a bounded interval in $\mathbb{R}$. Hence, without changing $V(X + tY)$, we can restrict v to some bounded interval and then extend v by periodicity to the real line. Since this modified v is three times differentiable, its Fourier coefficients decay rapidly; and we can establish differentiability of $V(X+tY)$ by estimating the derivatives of $\mathrm{trace}_n \exp(iu(X+tY))$. From Duhamel's formula follow the estimates required to show that V is twice continuously differentiable.

On a set $\Omega \subset M_n^s(\mathbb{R}^n)$ of full measure, the eigenvalues of X are distinct, and here we can calculate the successive partial derivatives of λ_j with respect to the matrix entries by using the perturbation theory of eigenvalues.[25] At $X \in \Omega$ we can write the Hessian of V in terms of the eigenvalues and the potential as

$$\langle \mathrm{Hess}\, V, Y \otimes Y \rangle = \frac{1}{n} \sum_{j=1}^{n} \{ v''(\lambda_j) \langle \nabla \lambda_j \mid Y \rangle^2 + v'(\lambda_j) \langle \mathrm{Hess}\, \lambda_j, Y \otimes Y \rangle \} \tag{6.3}$$

for each $Y \in M_n^s(\mathbb{R})$. This may be expressed more conveniently in terms of an orthonormal basis $(\xi_j)_{j=1}^n$ of eigenvectors of X corresponding to eigenvalues $(\lambda_j)_{j=1}^n$, with the Rayleigh–Schrödinger formula[25] giving

$$\begin{aligned} \langle \mathrm{Hess}\, V, Y \otimes Y \rangle = {} & \frac{1}{n} \sum_{j=1}^{n} v''(\lambda_j) \langle Y\xi_j, \xi_j \rangle_{\mathbb{R}^n}^2 \\ & + \frac{2}{n} \sum_{j,k:\, j<k} \frac{v'(\lambda_j) - v'(\lambda_k)}{\lambda_j - \lambda_k} \langle Y\xi_j, \xi_k \rangle_{\mathbb{R}^n}^2. \end{aligned} \tag{6.4}$$

It follows from (6.4) and the mean value theorem that V is convex whenever v is convex; and furthermore that V is uniformly convex in the sense of (6.2) whenever v is uniformly convex. We can extend the inequality (6.2) from the dense set Ω to all of $M_n^s(\mathbb{R})$ by continuity.

The condition (6.2) implies uniform 2-convexity in the sense of (1.8), and is analogous to the conditions (4b) and (23) of Bakry and Emery,[1] who establish hypercontractivity results for Markovian semigroups.

One can show either directly,[3] or by using the theory of logarithmic Sobolev inequalities,[1, 23] that such a V satisfies the following transportation inequality:

Lemma 6.3. Let V satisfy (6.2) for every n. Then all probability measures ω, which are absolutely continuous and of finite relative entropy with respect to ν_n, satisfy

$$W_2(\omega, \nu_n)^2 \leqslant \frac{2}{n^2 \delta} \mathrm{Ent}(\omega \mid \nu_n) \tag{6.5}$$

We remark that the helpful term $1/n^2$ in the constant of (6.5) originates from the n^2 in (6.1).

Proof of Theorem 6.1. Given Lemma 6.3, we can follow the same approach towards proving weak almost sure convergence of the empirical eigenvalue distribution as we followed in Sections 4 and 5. Let σ_n be the measure on $\mathbb{S}^n$ induced by the 1-Lipschitz eigenvalue map Λ. Then by duality σ_n satisfies a transportation inequality similar to (6.5).

Arguing as in the proof of Theorem 1.6(a), we see that any 1-Lipschitz function $F: (\mathbb{S}^n, \ell^2(n)) \to \mathbb{R}$ with $\int F(\lambda)\, \sigma_n(d\lambda) = 0$ satisfies the concentration inequality

$$\int_{\mathbb{S}^n} \exp\{tF(\lambda)\}\, \sigma_n(d\lambda) \leqslant \exp\left(\frac{t^2}{2n^2\delta}\right) \qquad (t \in \mathbb{R}). \tag{6.6}$$

Likewise σ_n satisfies the isoperimetric inequality

$$\sigma_n(A_\varepsilon) \geqslant 1 - \exp\left(-\frac{\delta n^2 \varepsilon^2}{8}\right) \qquad (\varepsilon > 0) \tag{6.7}$$

for all measurable $A \subset \mathbb{S}^n$ with $\sigma_n(A) \geqslant \frac{1}{2}$. We can now complete the proof of the Theorem 1.2 with the same technique as in Section 5, by applying a concentration inequality to $F(\lambda) = \frac{1}{n} \sum_{j=1}^{n} f(\lambda_j)$ for suitable choices of A.

7. PERTURBATION OF POTENTIALS

The methods of the preceding sections appear to be special to uniformly p-convex potentials $V(X)$, but one should expect some of the results to hold for certain multiple-well potentials obtained by perturbation. Let u be a uniformly bounded and continuously differentiable real function, and set $U(X) = \operatorname{trace}_n u(X)$ for $X \in M_n^s(\mathbb{R})$. We introduce various probability measures which are counterparts to those of section one: the ensemble

$$\nu_n^u(dX) = Z(n, u)^{-1} \exp\{-n^2(U(X) + V(X))\}\, dX \tag{7.1}$$

on $M_n^s(\mathbb{R})$, for normalizing constant $0 < Z(n, u) < \infty$; the corresponding eigenvalue distribution $\sigma_n^u(d\lambda)$ on $\mathbb{S}^n$; together with the empirical distribution under this law μ_n^u, and the associated integrated density of states χ_n^u on $\mathbb{R}$.

The ensemble ν_n^u of (7.1) may be regarded as a modification of ν_n of (1.4), and the strength of the perturbation is partly measured by the "free energy" expression[12]

$$F_{u,n}(t) = \log \int_{M_n^s(\mathbb{R})} \exp\{-tn^2 U(X)\}\, \nu_n(dX) \qquad (t \in \mathbb{R}). \tag{7.2}$$

Boutet de Monvel, Pastur and Shcherbina[7] show that, as $n \to \infty$, the μ_n^u converge weakly in probability to ρ^u, the weak limit of the integrated densities of states χ_n^u. They show further that ρ^u is the solution of a certain variational problem; when the potential $U+V$ has multiple wells, the support of ρ^u can consist of several disjoint intervals, in contrast to Proposition 1.3. See ref. 9. The following results give information as to the comparative positions of the supports of ρ and ρ^u.

Proposition 7.1. There exist $0 < t_n < 1$ such that

$$W_p^p(\chi_n^u, \chi_n) \leqslant \frac{C_p}{n^2} F''_{u,n}(t_n) \qquad (n \geqslant 1). \tag{7.3}$$

Proof. It follows from the Kantorovich–Rubinstein duality formula (1.16) and Lemma 4.1 that the transportation costs satisfy

$$W_p^p(\chi_n^u, \chi_n) \leqslant W_p^p(\nu_n^u, \nu_n). \tag{7.4}$$

The main transportation inequality (1.17) shows this to be bounded by the relative entropy

$$\leqslant \frac{C_p}{n^2} \operatorname{Ent}(\nu_n^u \mid \nu_n)$$

$$= \frac{C_p}{n^2} \left(\int_{M_n^s(\mathbb{R})} -n^2 U(X)\, \nu_n^u(dX) - F_{u,n}(1) \right). \tag{7.5}$$

We bound the final term by applying Jensen's inequality to (7.2) and obtain

$$W_p^p(\chi_n^u, \chi_n) \leqslant C_p \int_{M_n^s(\mathbb{R})} U(X)(\nu_n(dX) - \nu_n^u(dX)); \tag{7.6}$$

which may otherwise be expressed as

$$W_p^p(\chi_n^u, \chi_n) \leqslant \frac{C_p}{n^2}\,(F'_{u,n}(1) - F'_{u,n}(0)). \tag{7.7}$$

On applying the mean value theorem, we deduce the stated result.

Corollary 7.2. (a) For $p \geqslant 2$ even and $a_p > 0$, let $v(x) = a_p x^p$ and $u(x) = \sum_{j=0}^{p-1} a_j x^j$. Then the Wasserstein distance between the limiting integrated densities of states associated with the potentials $u+v$ and v satisfies

$$W_p(\rho^u, \rho)^p \leqslant \liminf_{n\to\infty} \frac{C_p}{n^2}\, F''_{u,n}(t_n). \tag{7.8}$$

(b) A similar result holds when $v(x) = a_p|x|^p$, with $2 \leqslant p < \infty$ and $a_p > 0$, and u is a uniformly bounded and continuously differentiable real function.

Proof. (a) By the triangle inequality

$$W_p(\rho^u, \rho) \leqslant W_p(\chi_n^u, \chi_n) + W_p(\chi_n, \rho) + W_p(\chi_n^u, \rho^u), \tag{7.9}$$

where the first summand may be bounded using Proposition 7.1. Such a u is Lipschitz on each bounded subset of $\mathbb{R}$. By Theorem 2 of Boutet de Monvel[7] *et al.*, χ_n^u converges weakly to ρ^u and χ_n converges weakly to ρ as $n \to \infty$. We shall improve this to convergence in the Wasserstein metric by showing that the sequences of measures $|x|^p \chi_n^u(dx)$ and $|x|^p \chi_n(dx)$ are uniformly tight. For the latter sequence, this follows from (4.14).

One checks that

$$\sup_{n\geqslant 1} \int_{[|x|\geqslant R]} |x|^p\, \chi_n^u(dx) = \sup_{n\geqslant 1} \int_{[\|X\|_{c^p(n)}\geqslant R]} \|X\|^p_{c^p(n)}\, \nu_n^u(dX), \tag{7.10}$$

and we require to show that these terms converge to zero as $R \to \infty$. It follows from the non-commutative Hölder inequality[26] that there exist constants κ_1 and κ_2 with

$$|U(X)| \leqslant \kappa_1 + \kappa_2\, \|X\|^{p-1}_{c^p(n)} \qquad (X \in M_n^s(\mathbb{R})). \tag{7.11}$$

Consequently we can apply Jensen's inequality to deduce

$$F_{u,n}(1) \geqslant -n^2 \int_{M_n^s(\mathbb{R})} (\kappa_1 + \kappa_2 \|X\|_{c^p(n)}^{p-1})\, \nu_n(dX) \geqslant -n^2 \kappa_3, \tag{7.12}$$

where the existence of this latest constant follows from (4.14). We can now write

$$(7.10) \leqslant \sup_{n \geqslant 1} \int_{[\|X\|_{c^p(n)} \geqslant R]} \exp\{n^2(\kappa_1 + \kappa_3 + \kappa_2 \|X\|_{c^p(n)}^{p-1})\}\, \nu_n(dX); \tag{7.13}$$

and the latest integral converges to zero as $R \to \infty$ on account of (4.14).

(b) The proof when u is uniformly bounded is similar but easier, for instead of (7.11) we have a constant M with $|U(X)| \leqslant M$ for all $X \in M_n^s(\mathbb{R})$.

8. CONCLUDING REMARKS

The results of the preceding sections and consideration of the Coulomb gas model of Wigner and Dyson[20] lead us to formulate the following:

Conjecture 8.1. Let w be a twice continuously differentiable real function for which there exists $\delta > 0$ with $w''(x) \geqslant \delta$ outside of some compact set. Then for $\beta \geqslant 0$, the ensemble

$$\varpi_n(d\lambda) = Z(n, w)^{-1} \prod_{j,k:\, j \neq k} |\lambda_j - \lambda_k|^\beta \cdot \exp\left(-n \sum_{j=1}^n w(\lambda_j)\right) d\lambda_1\, d\lambda_2 \cdots d\lambda_n \tag{8.1}$$

defines a probability measure on the simplex $\mathbb{S}^n$ for which Theorem 1.2 holds, so that the empirical distribution of eigenvalues $\mu_n = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j}$ under this law converges weakly almost surely to some non-random compactly supported probability measure on $\mathbb{R}$ as $n \to \infty$. Further, ϖ_n satisfies Theorems 1.4(b) and 1.5(b) with constants growing like n^2 as $n \to \infty$.

In support of this Conjecture, we consider the case of $\beta = 1$ and $w = u + v$, where u and v are as in Corollary 7.2(a). Using results of Bobkov, Ledoux[5] and Götze,[4] one can show that each of the measures satisfies the logarithmic Sobolev inequality

$$\int_{\mathbb{S}^n} f^2 \log f^2 / \|f\|_{L^2(\varpi_n)}^2 \, d\varpi_n \leqslant L_n(w) \int_{\mathbb{S}^n} \|\nabla f\|_{\mathbb{R}^n}^2 \, d\varpi_n \tag{8.2}$$

for all non-negative f in $L^2(\varpi_n)$ with $L^2(\varpi_n)$ distributional gradient. This leads to transportation inequalities $W_1(\alpha, \varpi_n)^2 \leqslant 2L_n(w)\,\mathrm{Ent}(\alpha \mid \varpi_n)$ by the results of Bobkov and Götze[4], and hence to variants of Theorem 1.4(b), 1.5(b) and 1.6(b). Unfortunately, our bounds on the constants $L_n(w)$ do not decay as $n \to \infty$, and we have not been able to obtain an appropriate version of Theorem 1.2. The said bounds use the perturbation estimates on logarithmic Sobolev constants, due to Deuschel[10] and Stroock, which are not well adapted to the present setting. Kac[17] obtains estimates on the heat kernel for killed Brownian motion. His results suggest that the logarithmic Sobolev inequality should hold for the measure ϖ_n, which is invariant under a diffusion process for the eigenvalues similar to that of the Coulomb gas model in ref. 20, p.195.

ACKNOWLEDGMENTS

This work was motivated by lectures at CIRM Luminy and Institute Henri Poincaré, Paris. I am grateful to M. Ledoux, J. Tawn, S. Thorbjørnsen and J. Groves for helpful conversations, and to the referees for improvements to the presentation.

REFERENCES

1. D. Bakry and M. Emery, Diffusions hypercontractives, in *Séminaire de Probabilités XIX*, J. Azéma et M. Yor, eds., Lecture Notes in Mathematics, Vol. 1123, pp. 177–206 (Springer, Berlin, 1985).
2. K. Ball, E. A. Carlen, and E. H. Lieb, Sharp uniform convexity and smoothness inequalities for trace norms, *Invent. Math.* **115**:463–482 (1994).
3. G. Blower, The Gaussian isoperimetric inequality and transportation, *Submitted to Positivity*.
4. S. G. Bobkov and F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Funct. Anal.* **163**:1–28 (1999).
5. S. G. Bobkov and M. Ledoux, From Brunn-Minkowski to Brescamp-Lieb and to logarithmic Sobolev inequalities, *Geom. Funct. Anal.* **10**:1028–1052 (2000).
6. C. Borell, The Brunn-Minkowski inequality in Gauss space, *Invent. Math.* **30**:207–216 (1975).
7. A. Boutet de Monvel, L. Pastur, and M. Shcherbina, On the statistical mechanics approach to random matrix theory: integrated density of states, *J. Stat. Phys.* **79**:585–611 (1995).
8. E. Brézin, C. Itzykson, G. Parisi, and J.B. Zuber, Planar diagrams, *Comm. Math. Phys.* **59**:35–81 (1978).
9. P. Deift, T. Kreicherbauer, K. T. R. McLaughlin, S. Venakides, and X. Zhou, Asymptotics for polynomials orthogonal with respect to varying exponential weights, *Internat. Math. Res. Notices* **16**:759–782 (1997).
10. J.-D. Deuschel and D. W. Stroock, Hypercontractivity and spectral gap of symmetric diffusions with applications to the stochastic Ising models, *J. Funct. Anal.* **92**:30–48 (1990).

10. P. Deift, T. Kreicherbauer, K. T. R. McLaughlin, S. Venakides, and X. Zhou, Asymptotics for polynomials orthogonal with respect to varying exponential weights, *Internat. Math. Res. Notices* **16**:759–782 (1997).
11. R. M. Dudley, *Real Analysis and Probability* (Wadsworth & Brooks/Cole, Pacific Grove, 1989).
12. R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics* (Springer, New York, 1985).
13. L. Gross, Logarithmic Sobolev inequalities and contractivity properties of semigroups, in *Lecture Notes in Math.*, Vol. 1563, pp. 55–88 (Springer, Berlin, 1993).
14. J. S. Groves, A study of Stochastic processes in Banach space, Ph.D. thesis (Lancaster University, 2000).
15. U. Haagerup and S. Thorbjørnsen, Random matrices with complex Gaussian entries, *Preprint*. (Odense, 1998).
16. K. Johansson, On fluctuations of eigenvalues of random Hermitian matrices, *Duke Math. J.* **91:**151–204 (1998).
17. M. Kac, On some connections between probability theory and differential and integral equations, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1951), pp. 189–215.
18. M. K.-H. Kiessling and H. Spohn, A note on the eigenvalue density of random matrices, *Comm. Math. Phys.* **199**:683–695 (1999).
19. M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, in *Séminaire de Probabilités XXXIII*, Lecture Notes in Math., Vol. 1709, pp. 120–216 (Springer, Berlin, 1999).
20. M. L. Mehta, *Random Matrices*, 2nd ed. (Academic Press, San Diego, 1991).
21. V. D. Milman and G. Schechtman, Asymptotic theory of finite dimensional normed spaces, Springer Lecture Notes in Math., Vol. 1200, (Springer, Berlin, 1986).
22. N. I. Muskhelishvili, *Singular Integral Equations: Boundary Problems of Function Theory and Their Application to Mathematical Physics* (P. Noordhoff N.V., Groningen, 1953).
23. F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* **173**:361–400 (2000).
24. G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry* (Cambridge University Press, 1989).
25. M. Reed and B. Simon, *Methods of Modern Mathematical Physics: Volume 4, Analysis of Operators* (Academic Press, New York, 1978).
26. B. Simon, *Trace Ideals and Their Applications* (Cambridge University Press, 1979).
27. M. Talagrand, Transportation cost for Gaussian and other product measures, *Geom. Funct. Anal.* **6**:587–600 (1996).
28. N. Tomczak-Jaegermann, The moduli of smoothness and convexity and the Rademacher averages of trace classes S_p $(1 \leqslant p < \infty)$ *Studia Math.* **50**:163–182 (1974).
29. N. Tomczak-Jaegermann, *Banach–Mazur Distances and Finite-Dimensional Operator Ideals* (Longman, Harlow Essex, 1989).