Bernoulli **10**(5), 2004, 755–782

Transportation of measure, Young diagrams and random matrices

GORDON BLOWER

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.
E-mail: g.blower@lancaster.ac.uk

The theory of transportation of measure for general convex cost functions is used to obtain a novel logarithmic Sobolev inequality for measures on phase spaces of high dimension and hence a concentration-of-measure inequality. There are applications to the Plancherel measure associated with the symmetric group, the distribution of Young diagrams partitioning N as $N \to \infty$ and to the mean-field theory of random matrices. For the potential $\log \Gamma(x+1)$, the generalized orthogonal ensemble and its empirical eigenvalue distribution are shown to satisfy a Gaussian concentration-of-measure phenomenon. Hence the empirical eigenvalue distribution converges weakly almost surely as the matrix size increases; the limiting density is given by the derivative of the Vershik probability density.

Keywords: infinite symmetric group; logarithmic Sobolev inequality; Young tableau

1. Introduction

This paper presents a new approach to transportation inequalities and the concentration-of-measure phenomenon for certain Gibbs measures on phase spaces of high dimension. The measures in question are not quite product measures, and the potentials are not uniformly convex. The applications include the following model problem of Dyson (1962). Let us take n positive unit charges and place them at points $\ell_1 > \ell_2 > \ldots > \ell_n > 0$ on the real line. The charges are mutually repelling and we suppose that they are subject to an electrostatic field with potential $v(x) = \log \Gamma(x+1)$, so that the total energy is

$$V(\ell) = \sum_{j=1}^{n} v(\ell_j) - \sum_{j,k=1: j<k}^{n} \beta \log(\ell_j - \ell_k). \tag{1.1}$$

Here $\beta > 0$ is a scale factor. By convexity of V there should exist some equilibrium configuration that minimizes the potential.

We regard the $\ell = (\ell_j)_{j=1}^n$ as points in a suitable phase space and $V(\ell)$ as the potential of a Gibbs probability measure $\nu(\mathrm{d}\ell) = Z^{-1} \mathrm{e}^{-V(\ell)} \mathrm{d}\ell_1 \ldots \mathrm{d}\ell_n$ involving the product Lebesgue measure. Henceforth we regard the ℓ as random subject to $\nu(\mathrm{d}\ell)$, and we use Z to stand for a typical partition function (normalizing constant).

The properties as $n \to \infty$ of such ensembles have previously been considered in two apparently disparate problems.

1. *Random permutations.* Let $N \geqslant 1$ be an integer and let $[N] = \{1, 2, \ldots, N\}$. The

1350–7265

group S_N of permutations of $[N]$ has a unique translation-invariant probability measure μ_N defined by $\mu_N(B) = \#(B)/N!$ and has a family of inequivalent irreducible representations that are parametrized by Young diagrams; see Fulton (1997). Such diagrams may be viewed as partitions of N, so that

$$\lambda \in \Omega_N := \left\{ (\lambda_j)_{j=1}^N \in \mathbb{Z}^N : \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_N \geqslant 0; \sum_{j=1}^N \lambda_j = N \right\}; \tag{1.2}$$

we often write $\lambda \vdash N$ and let $n = n(\lambda)$ be the largest index with $\lambda_n > 0$. We can represent the diagram as a left-justified array of square boxes with a row of λ_{j+1} boxes below a row of λ_j boxes for $j \geqslant 1$. A *standard Young tableau* is a Young diagram with its boxes numbered from $1, 2, \ldots, N$ so that all of the rows and columns are strictly increasing as one moves downwards or to the right.

There are f_λ distinct ways of numbering a Young diagram to give a standard Young tableau, where f_λ is given by the hook-length formula

$$f_\lambda = N! \frac{\prod_{j,k\,:1 \leqslant j < k \leqslant n} (\ell_j - \ell_k)}{\prod_{j=1}^n \ell_j!}, \qquad V(\ell) = -\log f_\lambda, \tag{1.3}$$

with integers $\ell_j = \lambda_j + n - j$ and $\beta = 1$. Up to isomorphism, all irreducible representations of S_N belong to this collection, so it follows from Frobenius's theorem that, with δ_λ the unit point mass on the partition λ,

$$\nu_N = \sum_{\lambda:\lambda \vdash N} \frac{f_\lambda^2}{N!} \delta_\lambda \tag{1.4}$$

defines a probability measure on the set of inequivalent irreducible representations of S_N; this is called the *Plancherel measure* of S_N.

The quantity λ_1 represents the length of the longest increasing subsequence of a randomly chosen permutation, and a famous problem of Ulam was to show that $\lambda_1/N^{1/2}$ converges in probability to 2 as $N \to \infty$. Vershik and Kerov (1977) showed, moreover, that the shape of the scaled Young diagrams converges in probability to the continuous distribution on $[0, 2]$ that is specified by (5.4) below. On account of the scaling, it is not important whether the λ_j are integers. By a variational argument in a suitable space of continuous functions, Logan and Shepp (1977) achieved a similar result. In this paper we do not seek to improve upon the combinatorial results that have been established for the discrete Plancherel measure as in Deuschel and Zeitouni (1999); rather in Section 4 we achieve new results for its continuous analogue.

2. *Random matrices*. Let $M_n^+(\mathbb{R})$ be the space of positive definite real symmetric matrices. Each $X \in M_n^+(\mathbb{R})$ has a unique list of eigenvalues $\lambda = (\lambda_j)_{j=1}^n$, in decreasing order according to multiplicity, which determines an element of the simplex

$$\Delta^n = \{(\lambda_j) \in \mathbb{R}_+^n : \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0\}.$$

In Section 6 we shall introduce a potential and Gibbs measure ν_n on the phase space $M_n^+(\mathbb{R})$ such that ν_n is invariant under the conjugation action $X \mapsto UXU^\dagger$ of the orthogonal group on $M_n^+(\mathbb{R})$; this is a variant of the generalized orthogonal ensemble of Dyson (1962). Further, with the potential $v(x) = \log\Gamma(x+1)$ and $\beta = 1$, the ordered eigenvalues have joint distribution

$$\omega_n(\mathrm{d}\lambda) = Z_n^{-1} \exp\left\{-\sum_{j=1}^n v(\lambda_j)\right\} \prod_{j,k=1:j<k}^n (\lambda_j - \lambda_k)^\beta \,\mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_n. \tag{1.5}$$

While there is a well-developed analogy between the distributions of row lengths of Young diagrams and the eigenvalues of Gaussian random matrices, as presented in Baik *et al*. (1999) and Borodin *et al*. (2000), the methods used in the two settings have often appeared rather different and unrelated to the general concentration-of-measure phenomenon in phase space of high dimension. Transportation inequalities in the style of Talagrand (1996) provide a link between the two theories. Blower (2001) proved concentration inequalities for eigenvalues of random matrices under the generalized orthogonal ensemble for potentials such as $v(x) = x^2$ that are uniformly convex; in this paper the method is extended to potentials that are not uniformly convex. To deal with this we establish new transportation inequalities for a general class of convex cost function.

In Section 2 we recall basic results about transportation of measure which will be used in subsequent sections. In Section 4 we consider the convexity properties of V of (1.1), and deduce logarithmic Sobolev inequalities and concentration inequalities for the continuous Gibbs measure. These results are deduced from an abstract logarithmic Sobolev inequality for general convex cost functions that is presented in Section 3. The proof of Theorem 3.1, which extends that of Bobkov and Ledoux (2000), depends upon the following inequality, due to Prékopa and Leindler. Let f, g and h be positive and measurable functions on $\mathbb{R}^n$ such that $f(sx + (1-s)y) \geqslant g(x)^s h(y)^{1-s}$ for some $0 < s < 1$ and all $x, y \in \mathbb{R}^n$. Then

$$\int_{\mathbb{R}^n} f(z)\mathrm{d}z \geqslant \left(\int_{\mathbb{R}^n} g(x)\mathrm{d}x\right)^s \left(\int_{\mathbb{R}^n} h(y)\mathrm{d}y\right)^{1-s}. \tag{1.6}$$

Pisier (1989, Lemma 1.2) presents a simple proof.

In Section 5 we consider the empirical distribution of row lengths of Young diagrams as $N \to \infty$. The empirical distribution has similar properties to the probability measure $(1/n)\sum_{j=1}^n \delta_{\lambda_j/n}$, where $(1/n)\#\{j : \lambda_j/n \leqslant x\}$ is often called the eigenvalue counting function. Then in Section 6 we show how the probability measure (1.5) does indeed arise for the joint eigenvalue distributions of Dyson's generalized orthogonal ensemble with potential function $\log\Gamma(1+x)$. In Section 7 we use the concentration inequalities of Section 6 to obtain convergence as $n \to \infty$ for various correlation functions associated with the random matrices. A significant virtue of mean-field and transportation methods is that they apply at the level of joint eigenvalue distributions for all $\beta > 0$, and hence apply likewise to orthogonal ensembles of real symmetric matrices with $\beta = 1$, unitary ensembles of complex Hermitian matrices with $\beta = 2$, and symplectic ensembles of quaternion matrices with $\beta = 4$. We focus on orthogonal ensembles as Johansson's (1998) theory covers unitary ensembles, where the technique of orthogonal polynomials is effective and more precise results are known.

2. Induced measures and transportation

Let (Ω_j, d_j) $(j = 1, 2)$ be compact metric spaces. We say that $f : \Omega_1 \to \Omega_2$ is an *L-Lipschitz* function if $d_2(f(x_1), f(x_2)) \leqslant L d_1(x_1, x_2)$ for all $x_1, x_2 \in \Omega_1$ and some L.

Given a continuous map $\Phi : \Omega_1 \to \Omega_2$, we say that Φ *induces* the Radon measure ϖ_2 on Ω_2 from the Radon measure ϖ_1 on Ω_1 when

$$\int_{\Omega_2} f(y)\varpi_2(\mathrm{d}y) = \int_{\Omega_1} f(\Phi(x))\varpi_1(\mathrm{d}x) \tag{2.1}$$

holds for all continuous $f : \Omega_2 \to \mathbb{R}$.

By a *cost function* we mean a continuous function $c : \Omega_1 \times \Omega_1 \to [0, \infty)$ with $c(x, y) = c(y, x)$ and $c(x, x) = 0$ for all $x, y \in \Omega_1$; evidently $c(x, y) = d(x, y)^p$ gives an example for $1 \leqslant p$. Given Radon probability measures w_1 and w_2 on Ω_1, the *transportation cost* of taking w_1 to w_2 with respect to the cost function c is

$$\mathrm{Tc}_c(w_1, w_2) = \inf_{\pi}\left\{\iint_{\Omega_1\times\Omega_1} c(x, y)\pi(\mathrm{d}x\,\mathrm{d}y)\right\} \tag{2.2}$$

where π is a probability measure on $\Omega_1 \times \Omega_1$ with marginals w_1 and w_2. The Kantorovich–Rubinstein duality formula asserts that

$$\mathrm{Tc}_c(w_1, w_2) = \sup_{f,g}\left\{\int_{\Omega_1} f(x)w_1(\mathrm{d}x) - \int_{\Omega_1} g(y)w_2(\mathrm{d}y) : f(x) - g(y) \leqslant c(x, y)\right\}, \tag{2.3}$$

where $f, g : \Omega_1 \to \mathbb{R}$ are continuous and bounded functions. When $c(x, y) = \|x - y\|^p_{\mathbb{R}^n}$ for $x, y \in \mathbb{R}^n$, we write Tc_p for the transportation cost and note that $\mathrm{T}c_1$ gives the *Wasserstein metric* on probability measures. See Gangbo and McCann (1996) for a historical survey and recent developments on this topic.

When w_1 is absolutely continuous with respect to w_2, we define the *relative entropy* of w_1 with respect to w_2 by

$$\mathrm{Ent}(w_1|w_2) = \int_{\Omega_1} \frac{\mathrm{d}w_1}{\mathrm{d}w_2} \log \frac{\mathrm{d}w_1}{\mathrm{d}w_2}\, \mathrm{d}w_2, \tag{2.4}$$

where by Jensen's inequality $0 \leqslant \mathrm{Ent}(w_1|w_2) \leqslant \infty$.

In the next section we prove a transportation inequality which bounds the transportation cost $\mathrm{Tc}_c(w_1, w_2)$ by $\mathrm{Ent}(w_1|w_2)$ for a suitable cost function c, a probability measure w_2 with special properties in relation to c, and a general w_1.

3. Logarithmic Sobolev inequality for convex cost functions

Let I be an open interval and c a convex cost function on I^2, so that:

(i) $c(x, y) > 0$ for $x \neq y$, and $c(x, x) = 0$;
(ii) $c(x, y) = c(y, x)$, for all $x, y \in \mathrm{I}$;
(iii) c is continuously differentiable on I^2;
(iv) c is convex on I^2;
(v) $c(x, y) \leqslant a(x) + b(y)$ for some continuous functions $a, b : I \to \mathbb{R}$.

To extend the cost function to n variables, we introduce $c_{(n)}(x, y) = \sum_{j=1}^{n} c(x_j, y_j)$ for $x = (x_j)$, $y = (y_j) \in I^n$; we shall usually suppress the subscript n for simplicity.

We define the *dual cost function* for suitable $\xi = (\xi_j) \in \mathbb{R}^n$ by

$$c^*(x, \xi) = \sup_y \{\langle y - x, \xi\rangle - c(x, y) : y \in I^n\}, \qquad x \in I^n, \tag{3.1}$$

with $\langle\cdot, \cdot\rangle$ the standard inner product; c^* may be regarded as the Legendre transform of $c(x, y)$ in the second variable. Hence $c^*(x, \xi)$ is convex in ξ, and $(x \mapsto -c^*(x, \xi) - \langle x, \xi\rangle)$ is c-concave in the sense of Gangbo and McCann (1996); see also Villani (2003, p. 33).

The dual cost function $c^*(x, \xi)$ is finite whenever ξ belongs to $\{\nabla_2 c(x, z) : z \in I^n\}$; here ∇_2 represents the gradient in the second variable. Indeed, by convexity we have

$$c(x, y) \geqslant c(x, z) + \langle\nabla_2 c(x, z), y - z\rangle,$$

where we let $\xi = \nabla_2 c(x, z)$, which is independent of y; hence

$$c^*(x, \xi) = \sup_y \{\langle y - x, \xi\rangle - c(x, y)\} \leqslant \langle z - x, \xi\rangle - c(x, z) < \infty.$$

One can recover c from c^* by the formula $c(x, y) = \sup_\xi\{\langle y - x, \xi\rangle - c^*(x, \xi) : c^*(x, \xi) < \infty\}$.

We shall further assume that:

(vi) $c^*(x, \xi)$ is finite and continuous for $x \in I^n$ and $\xi \in [-1, 1]^n$.

The principal examples of cost functions in this paper are as follows.

Example 3.1. Let Φ be a continuously differentiable, strictly increasing, and convex function on $[0, \infty)$ with $\Phi(0) = 0$. Then $c(x, y) = \Phi(|x - y|)$ gives a translation-invariant cost function for $x, y \in \mathbb{R}$ such that $c^*(x, \xi) = \sup\{\xi z - \Phi(|z|) : z \in \mathbb{R}\}$ is the usual Legendre transform of $\Phi(|z|)$. This case is thoroughly discussed in the appendices of Gangbo and McCann (1996). The choice of $\Phi(x) = x^2/2$ gives the quadratic cost function, for which $c^*(x, \xi) = \xi^2/2$, as in Talagrand (1996) and Otto and Villani (2000).

Example 3.2. With subsequent applications to (1.1) in mind, we let

$$c(x, y) = \frac{(x - y)^2}{x + y}, \qquad x, y > 0. \tag{3.2}$$

One can check by calculus that c is convex on $(0, \infty)^2$ and, moreover, that

$$c^*(x, u) = \begin{cases} \infty, & x > 0,\ u > 1, \\ 2x\{1-(1-u)^{1/2}\}^2, & x > 0,\ -3 \leqslant u \leqslant 1, \\ -x(u+1), & x > 0,\ u \leqslant -3. \end{cases} \tag{3.3}$$

To obtain this formula, note that when $u > 1$, the function

$$g(y) = u(y-x) - \frac{(y-x)^2}{y+x}$$

diverges to infinity as $y \to \infty$. Now take $-3 < u < 1$ and set $v = (y-x)/(y+x)$, which satisfies $-1 < v < 1$ for $0 < y < \infty$. The maximum of g occurs where $v = 1-(1-u)^{1/2}$ since $0 = g'(y) = u - 2v + v^2$. When $u < -3$, this stationary point also lies outside the acceptable range and the maximum occurs at $v = -1$; that is, $y = 0$.

This cost function does not satisfy the superlinear growth condition (H3) of Gangbo and McCann (1996), and hence their optimal transportation theory does not apply. Nevertheless, $c(x, y)$ has linear growth as $y \to \infty$ for fixed x, so (v) holds, and $c(x, y)$ behaves for x close to y like a scaled version of the quadratic cost function. These properties make it especially suitable for dealing with the potential (1.1) where the growth is like $y \log y$, just faster than linear. In Section 4 we consider this in detail.

In general, we shall exploit the linkage between potentials and cost functions c as above that is expressed in the following definition.

Definition. *Let $W : \Omega \to \mathbb{R}$ be a continuous potential function where Ω is a convex and open subset of I^n. We say that W is c-convex with constant $\kappa > 0$ if*

$$(1-s)W(x) + sW(y) - W((1-s)x + sy) \geqslant \kappa s(1-s)c(x, y) \tag{3.4}$$

holds for all $0 < s < 1$ and all $x, y \in \Omega$. Any c-convex potential function is strictly convex.

The following theorem generalizes a result of Schmuckenschläger presented by Bobkov and Ledoux (2000). Classical logarithmic Sobolev inequalities involve the quadratic cost function.

Theorem 3.1. *Let $\nu(\mathrm{d}x) = Z^{-1}\mathrm{e}^{-w(x)}\mathrm{d}x$ be the probability measure with potential W on Ω, where W is c-convex with constant κ and where c satisfies* (i)–(vi). *Then any positive function $f \in C^\infty(\Omega)$, with $|\partial f/\partial x_j| \leqslant \kappa f(x)$ for all $x \in \Omega$, satisfies the logarithmic Sobolev inequality*

$$\int_\Omega f(x) \log\left(f(x) \Big/ \int_\Omega f\,\mathrm{d}\nu\right)\nu(\mathrm{d}x) \leqslant \kappa \int_\Omega c^*\left(x, \frac{-\nabla f(x)}{\kappa f(x)}\right) f(x)\nu(\mathrm{d}x). \tag{3.5}$$

Proof. We may assume that $f(x) = \mathrm{e}^{k(x)}$, where k is continuously differentiable. Adjusting the normalizing constants if necessary, we can replace Ω by a compact convex subset Ω_c of $\mathbb{R}^n$, so that k has bounded derivatives of all orders on Ω_c.

By the Prékopa–Leindler inequality (1.6), if g_s is a function such that

$$g_s(z) - W(z) \geqslant k(x) - (1-s)W(x) - sW(y) \tag{3.6}$$

holds for some $s \in (0, 1)$ and all $x, y \in \Omega_c$ with $z = (1-s)x + sy$, then

$$Z^{-1}\int_{\Omega_c} \exp\{g_s(x) - W(x)\}\mathrm{d}x \geqslant \left(Z^{-1}\int_{\Omega_c} \exp\{(k(x)/(1-s)) - W(x)\}\mathrm{d}x\right)^{1-s} \tag{3.7}$$

also holds. To satisfy (3.6) for small s > 0, we select

$$g_s(z) = k(z) + \kappa s(1-s)c^*(z, -\nabla k(z)/\kappa(1-s)^2) + \alpha s^2, \tag{3.8}$$

where α is to be chosen so that g_s satisfies

$$g_s(z) + \kappa s(1-s)c(x, y) \geqslant k(x), \qquad x, y \in \Omega_c; \tag{3.9}$$

since W is c-convex, (3.9) implies (3.6). By the mean value theorem we have $k(z) = k(x) + \langle \nabla k(z), z - x\rangle + \langle \mathrm{Hess}\, k_{x_0}(z-x), z-x\rangle/2$ for some x_0 between x and z, where the final term, involving the Hessian of second-order partial derivatives, is $\mathrm{O}(\|z-x\|^2) = \mathrm{O}(s^2)$ as $s \to 0+$. Here the implied constants in O terms depend upon the diameter of Ω_c and the supremum on Ω_c of the norm of Hess k; hence they can be chosen uniformly for $x, y \in \Omega_c$.

So we shall have (3.9) once we verify

$$\begin{aligned}\kappa s(1-s)c^*\left(z, \frac{-\nabla k(z)}{\kappa(1-s)^2}\right) + \alpha s^2 &\geqslant \langle z - x, -\nabla k(z)\rangle - \kappa s(1-s)c(x, y)\\ &= \kappa s(1-s)\left(\left\langle y - z, \frac{-\nabla k(z)}{\kappa(1-s)^2}\right\rangle - c(z, y)\right)\\ &\quad + \kappa s(1-s)(c(z, y) - c(x, y)).\end{aligned} \tag{3.10}$$

By the mean value theorem there exists $\bar{x}$ between x and z such that $c(z, y) - c(x, y) = \langle z - x, \nabla_1 c(\bar{x}, y)\rangle$, which is of order $O(s)$, and the final term in (3.10) is $O(s^2)$; so we can use the definition (3.1) of c^* and take an appropriate α in (3.8) to satisfy (3.9). The implied constants in the O terms depend upon the diameter of Ω_c and the supremum of $\|\nabla_1 c(u, y)\|$ for u between x and z; hence they can be chosen uniformly for $x, y \in \Omega_c$.

We return to (3.7) and expand both sides as power series in s, thus obtaining

$$\begin{aligned}&Z^{-1}\int_{\Omega_c} \{1 + \kappa s(1-s)c^*(z, -\nabla k(z)/\kappa(1-s)^2) + O(s^2)\}\mathrm{e}^{k(z) - W(z)}\,\mathrm{d}z\\ &\geqslant \exp\left\{(1-s)\log\left(Z^{-1}\int_{\Omega_c}(1 + sk(x) + O(s^2))\mathrm{e}^{k(x)-W(x)}\mathrm{d}x\right)\right\}.\end{aligned} \tag{3.11}$$

At $s = 0$, both sides are equal; so the coefficients of s must respect the inequality, and we deduce

$$\kappa Z^{-1}\int_{\Omega_c} c^*(z, -\nabla k(z)/\kappa)\mathrm{e}^{k(z)-W(z)}\,\mathrm{d}z \tag{3.12}$$
$$\geqslant Z^{-1}\int_{\Omega_c} k(x)\mathrm{e}^{k(x)-W(x)}\mathrm{d}x - Z^{-1}\int_{\Omega_c}\mathrm{e}^{k(x)-W(x)}\,\mathrm{d}x\log\left(Z^{-1}\int_{\Omega_c}\mathrm{e}^{k(x)-W(x)}\,\mathrm{d}x\right),$$

an expression which is equivalent to (3.5). □

The following results are generalizations of results from Bobkov *et al*. (2001). The novelty lies in the wider choice of cost function; in particular, our cost functions need not be translation-invariant.

Theorem 3.2. *Let W, c and ν be as in Theorem 3.1. Then ν satisfies the transportation inequality*

$$\mathrm{Tc}_c(\mu, \nu) \leqslant \kappa^{-1}\,\mathrm{Ent}(\mu|\nu) \tag{3.13}$$

for all probability measures μ that are absolutely continuous and of finite relative entropy with respect to ν.

Proof. Suppose that $f, g : \Omega \to \mathbb{R}$ are continuous and bounded functions such that $g(y) - f(x) \leqslant c(x, y)$ for all $x, y \in \Omega$. Then

$$s((1-s)\kappa g(y) - W(y)) + (1-s)(-s\kappa f(x) - W(x)) \leqslant -W((1-s)x + sy) \tag{3.14}$$

holds for all $x, y \in \Omega$ and $0 < s < 1$ since W is c-convex; hence by the Prékopa–Leindler inequality (1.6) we have

$$\left(\int_\Omega \mathrm{e}^{(1-s)\kappa g(x)-W(x)}\,\mathrm{d}x/Z\right)^s\left(\int_\Omega \mathrm{e}^{-s\kappa f(x)-W(x)}\,\mathrm{d}x/Z\right)^{1-s} \leqslant \int_\Omega \mathrm{e}^{-w(x)}\,\mathrm{d}x/Z. \tag{3.15}$$

At $s = 0$, both sides equal one, so the right derivative of the left-hand side at $s = 0$ must be less than or equal to zero, hence

$$\log\left(\int_\Omega \mathrm{e}^{\kappa g(x)-W(x)}\,\mathrm{d}x/Z\right) - \kappa\int_\Omega f(x)\mathrm{e}^{-w(x)}\,\mathrm{d}x/Z \leqslant 0. \tag{3.16}$$

Theorem 63 of Hardy *et al*. (1952) gives rise to a dual formula for relative entropy,

$$\mathrm{Ent}(\mu|\nu) = \sup_h\left\{\int_\Omega h(y)\mu(\mathrm{d}y) : \int_\Omega \mathrm{e}^{h(x)}\nu(\mathrm{d}x) \leqslant 1\right\}, \tag{3.17}$$

in which we can take, on account of (3.16), the function $h(y) = \kappa\{g(y) - \int_\Omega f(x)\nu(\mathrm{d}x)\}$ and deduce that

$$\mathrm{Ent}(\mu|\nu) \geqslant \kappa\left(\int_\Omega g(y)\mu(\mathrm{d}y) - \int_\Omega f(x)\nu(\mathrm{d}x)\right). \tag{3.18}$$

The required result now follows from the Kantorovich–Rubinstein formula (2.3), where condition (v) serves as a substitute for compactness. Formulae similar to (3.18) appear in several papers, including Bobkov and Götze (1999). □

The dual form of Theorem 3.2 is the following concentration inequality.

Corollary 3.3. *Let W, c and ν be as in Theorem 3.1, and suppose that* $e^{\kappa\|x\|_1}$ *is integrable with respect to ν, where* $\|x\|_1 = \sum_{j=1}^n |x_j|$.

(i) *Then for any differentiable function g with $|\partial g/\partial x_j| \leqslant L$ for all $x \in \Omega$, the function*

$$Q_L(g(x)) = \inf_y \{Lc(x, y) + g(y) : y \in \Omega\} \tag{3.19}$$

is finite on Ω and satisfies

$$\int_\Omega \exp\{\kappa Q_L(g(x))/L\}\nu(\mathrm{d}x) \leqslant \exp\left\{\kappa\int_\Omega g(y)\nu(\mathrm{d}y)/L\right\}. \tag{3.20}$$

(ii) *If, moreover, g is convex, then*

$$\int_\Omega \exp\{(\kappa g(x)/L) - \kappa c^*(x, -\nabla g(x)/L)\}\nu(\mathrm{d}x) \leqslant \exp\left\{\kappa\int_\Omega g(y)\nu(\mathrm{d}y)/L\right\}. \tag{3.21}$$

Proof. (i) By the mean value theorem we can write

$$g(y) + Lc(x, y) = g(x) - L\{\langle y - x, \xi\rangle - c(x, y)\}, \tag{3.22}$$

where $\xi = -\nabla g(\bar{x})/L \in [-1, 1]^n$, and so $Q_L(g(x)) > -\infty$ by (vi). We observe that $\kappa g(x)/L$ grows more slowly than $\kappa\|x\|_1$ as $\|x\|_1 \to \infty$ and hence $e^{\kappa g(x)/L}$ is integrable with respect to $\nu(\mathrm{d}x)$. Further, $Q_L(g(x)) \leqslant g(x)$. As both sides of (3.20) define increasing functionals of g, we assume without loss that g is bounded above in the following computation, for one can then relax the bound with the aid of the monotone convergence theorem. For $0 < t < \kappa/L$ we introduce a normalizing constant Z_t such that

$$\mu_t(\mathrm{d}x) = Z_t^{-1} \exp\left(tQ_L(g(x)) - t\int_\Omega g(y)\nu(\mathrm{d}y)\right)\nu(\mathrm{d}x) \tag{3.23}$$

defines a probability measure. The derivative of Z_t satisfies

$$\begin{aligned}\frac{\mathrm{d}Z_t}{\mathrm{d}t} &= \int_\Omega \left(Q_L(g(x)) - \int_\Omega g\,\mathrm{d}\nu\right)\exp\left(tQ_L(g(x)) - t\int_\Omega g\,\mathrm{d}\nu\right)\nu(\mathrm{d}x)\\ &= Z_t\left(\int_\Omega Q_L(g(x))\mu_t(\mathrm{d}x) - \int_\Omega g(y)\nu(\mathrm{d}y)\right).\end{aligned} \tag{3.24}$$

We now use the Kantorovich–Rubinstein duality formula (2.3) and (3.19) to deduce that

$$\frac{\mathrm{d}Z_t}{\mathrm{d}t} \leqslant LZ_t\mathrm{Tc}_c(\mu_t, \nu) \tag{3.25}$$

and then Theorem 3.2 to obtain

$$\frac{1}{Z_t}\frac{\mathrm{d}Z_t}{\mathrm{d}t} \leqslant \frac{L}{\kappa}\int_\Omega \log\frac{\mathrm{d}\mu_t}{\mathrm{d}\nu}\,\mathrm{d}\mu_t. \tag{3.26}$$

From (3.23) and (3.26) we obtain the differential inequality

$$\frac{1}{Z_t}\frac{\mathrm{d}Z_t}{\mathrm{d}t} \leqslant \frac{L}{\kappa}\int_\Omega \left(-\log Z_t + tQ_L(g(x)) - t\int_\Omega g\,\mathrm{d}\nu\right)\mu_t(\mathrm{d}x)$$
$$= -\frac{L}{\kappa}\log Z_t + \frac{Lt}{\kappa Z_t}\frac{\mathrm{d}Z_t}{\mathrm{d}t}. \tag{3.27}$$

It is easy to integrate this differential inequality to obtain

$$(1 - Lt/\kappa)^{-1}\log Z_t - \log Z_0 \leqslant 0. \tag{3.28}$$

Hence $Z_t \leqslant 1$ holds for $0 < \mathrm{t} < \kappa/L$, since $Z_0 = 1$. Applying Fatou's lemma, we deduce (3.20) in the case of $t = \kappa/L$, as required.

(ii) When g is convex we have, as in (3.22),

$$Q_L(g(x)) \geqslant g(x) - Lc^*(x, -\nabla g(x)/L). \tag{3.29}$$

Inequality (3.21) is an immediate consequence of (3.20) and (3.29). □

Remarks. (i) The form of Q_L is suggested by the Hopf–Lax solution of the Hamilton–Jacobi equation; see Villani (2003, p. 175).

(ii) Marton showed that transportation inequalities may be converted into isoperimetric inequalities; see Talagrand (1996) or Blower (2001, Theorem 1.5) for details.

4. Logarithmic Sobolev inequality for Plancherel measure

In this section we consider the potential W that gives rise to (1.5), and present the results of Section 3 in a simple form for the associated Gibbs measure ω_n. This ω_n does not satisfy a logarithmic Sobolev inequality in the classical sense of quadratic cost functions. Whereas there is interaction between the eigenvalues in the potential (1.1), so ω_n is not a product measure, W is nevertheless c-convex, where c is a sum of suitable cost functions in the various directions.

Lemma 4.1. *For $\beta > 0$, the potential*

$$W(x) = \sum_{j=1}^{n} \log\Gamma(x_j) - \sum_{j,k:1\leqslant j<k\leqslant n} \beta\log(x_j - x_k) \tag{4.1}$$

on Δ^n is c-convex with constant $\kappa = \frac{1}{4}$ for the cost function

$$c(x, y) = \sum_{j=1}^{n}\frac{(x_j - y_j)^2}{x_j + y_j}, \qquad x_j, y_j > 0. \tag{4.2}$$

Proof. We observe that

$$W(x)+W(y)-2W\Big(\frac{x+y}{2}\Big)=\sum_{j=1}^{n}\Big\{\log\Gamma(x_j)+\log\Gamma(y_j)-2\log\Gamma\Big(\frac{x_j+y_j}{2}\Big)\Big\}$$
$$+\sum_{j,k:1\leqslant j<k\leqslant n}\beta\log\frac{(x_j-x_k+y_j-y_k)^2}{4(x_j-x_k)(y_j-y_k)}, \tag{4.3}$$

where the final sum is positive since $x_j-x_k>0$, $y_j-y_k>0$. To deal with a typical term in the first sum, we take $t=(x_j+y_j)/2$ and $h=(x_j-y_j)/2$, then apply the second mean value theorem to obtain

$$\log\Gamma(t+h)+\log\Gamma(t-h)-2\log\Gamma(t)=h^2\frac{\mathrm{d}^2}{\mathrm{d}t^2}\log\Gamma(\bar{t}),$$

for some $\bar{t}$ between the positive numbers $t-h$ and $t+h$. It follows from Euler's product formula (Copson, 1935) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\Gamma(t)=-\gamma+\sum_{k=1}^{\infty}\Big(\frac{1}{k}-\frac{1}{k+t-1}\Big) \tag{4.4}$$

and hence

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\log\Gamma(t)=\sum_{k=1}^{\infty}\frac{1}{(k+t-1)^2}\geqslant\frac{1}{t}. \tag{4.5}$$

From (4.3) follows the inequality

$$W(x)+W(y)-2W\Big(\frac{x+y}{2}\Big)\geqslant\frac{1}{2}\sum_{j=1}^{n}\frac{(x_j-y_j)^2}{x_j+y_j}=\frac{1}{2}c(x,\,y), \tag{4.6}$$

and this suffices to prove the result. □

Theorem 4.2. *Let* $\omega_n=Z_n^{-1}\mathrm{e}^{-W(n\lambda)}\,\mathrm{d}\lambda$ *be the probability measure on* Δ^n *that has scaled potential*

$$W(n\lambda)=\sum_{j=1}^{n}\log\Gamma(n\lambda_j)-\sum_{j,k\,:1\leqslant j<k\leqslant n}\beta\log(\lambda_j-\lambda_k). \tag{4.7}$$

Then any positive and continuously differentiable function f, such that $|\partial f/\partial x_j|\leqslant(n/4)f(x)$ *for all* $x\in\Delta^n$ *and* $j=1,\ldots,n$, *satisfies the logarithmic Sobolev inequality*

$$\int_{\Delta^n}f(x)\log\Big(f(x)\Big/\int_{\Delta^n}f\,\mathrm{d}\omega_n\Big)\omega_n(\mathrm{d}x)\leqslant\frac{8}{n}\int_{\Delta^n}\sum_{j=1}^{n}\frac{x_j}{f(x)}\Big(\frac{\partial f}{\partial x_j}\Big)^2\omega_n(\mathrm{d}x). \tag{4.8}$$

Proof. This is a special case of Theorem 3.1 for the cost function c introduced in Lemma 4.1. The corresponding c^* is given by (3.3) and satisfies the simple inequality $c^*(x,\,u)\leqslant 2xu^2$ for $x>0$ and $-1\leqslant u\leqslant 1$ in one dimension. On Δ^n we have

$$c^*(x, \xi) = \sum_{j=1}^{n} 2x_j \Big\{1 - (1 - \xi_j)^{1/2}\Big\}^2 \leqslant \sum_{j=1}^{n} 2x_j \xi_j{}^2, \tag{4.9}$$

where $\xi = -4(\nabla f)/(nf) \in [-1, 1]^n$. □

We can also obtain concentration inequalities that have a Gaussian form by simplifying the dual cost function.

Proposition 4.3. *Suppose that g is a continuously differentiable function on Δ_K^n, where $\Delta_K^n = \Delta^n \cap [0, K]^n$. Suppose further that $|\partial g/\partial x_j| \leqslant L$ and that $\int_{\Delta_K^n} g(x)\omega_n(\mathrm{d}x) = 0$. Then*

$$\omega_n\{x \in \Delta_K^n : |g(x)| \geqslant \varepsilon\} \leqslant 2\exp\{-\varepsilon^2/(32KL^2)\}, \qquad 0 < \varepsilon < 4KnL. \tag{4.10}$$

Proof. For the scaled potential of (4.7) we have $\kappa = n/4$. For $0 \leqslant t \leqslant 1$ and $x \in \Delta_K^n$ there exists $y \in \Delta_K^n$ such that $Q_L(tg(x)) = tg(x) - Lc^*(x, -t\nabla g(y)/L)$, and by (4.9) this gives

$$\begin{aligned} Q_L(tg(x)) &\geqslant tg(x) - 2L^{-1}t^2 \sum_{j=1}^{n} x_j \left|\frac{\partial g}{\partial x_j}\right|^2 \\ &\geqslant tg(x) - 2KLnt^2. \end{aligned} \tag{4.11}$$

It follows from Corollary 3.3 that

$$\int_{\Delta_K^n} \mathrm{e}^{\kappa Q_L(tg(x))/L} \omega_n(\mathrm{d}x) \leqslant 1, \tag{4.12}$$

and hence from (4.12), as in Chebyshev's inequality, that

$$\omega_n\{x \in \Delta_K^n : g(x) > 2KLnt + \eta L/(t\kappa)\} \leqslant \mathrm{e}^{-\eta}, \qquad \eta > 0. \tag{4.13}$$

When $0 < \varepsilon < 4KnL$ we can optimize this inequality by selecting $\eta = \varepsilon^2/(32KL^2)$ and $t = \varepsilon/(4KnL) < 1$. A similar argument works with $-g$, and we can deduce the required result (4.10). □

Corollary 4.4. *The probability measure ω_n on Δ^n satisfies*

$$\int_{\Delta^n} \exp\left(\frac{(4.2^{1/2} - 5)n}{4} \sum_{j=1}^{n} x_j\right) \omega_n(\mathrm{d}x) \leqslant \exp\left(\frac{n}{4} \int_{\Delta^n} \sum_{j=1}^{n} x_j \omega_n(\mathrm{d}x)\right). \tag{4.14}$$

Proof. This is an immediate consequence of Corollary 3.3(ii). By Lemma 1 of Boutet de Monvel *et al*. (1995), $\exp\{s\|x\|_1\}$ is ω_n-integrable for all $s \in \mathbb{R}$. □

5. The RSK correspondence and the Vershik distribution

We recall from Fulton (1997) the Robinson–Shensted–Knuth correspondence. There is a natural bijection between S_N and the set of pairs of standard Young tableaux with equal shape $\lambda \vdash N$, so we can form a map $\Phi : S_N \to \Omega_N : \sigma \leftrightarrow (P, Q) \mapsto \lambda(\sigma)$ that induces the Plancherel measure ν_N from the Haar measure μ_N.

We rotate and scale each Young diagram $\lambda \vdash N$ and associate it with the probability density function

$$h_\lambda(x) = \frac{1}{N^{1/2}} \sum_{j=1}^{n} \lambda_j \mathbb{I}_{[(j-1)N^{-1/2}, jN^{-1/2}]}(x), \tag{5.1}$$

where $\mathbb{I}_I$ stands for the indicator function of a set I.

The mean shape of a Young diagram with N boxes is represented by the probability density function

$$p_N(x) = \int_{S_N} h_{\lambda(\sigma)}(x)\mu_N(\mathrm{d}\sigma), \qquad x \in [0, \infty). \tag{5.2}$$

Now p_N is a decreasing function on $[0, \infty)$ for each N, and so by Helly's selection principle there exists a subsequence $(p_{N(k)})$ such that $p_{N(k)}(x) \to p_\Omega(x)$ as $N(k) \to \infty$. In fact, p_Ω is the probability density function on $[0, 2]$ of the *Vershik* Ω *distribution*, as in Vershik and Kerov (1977). To describe p_Ω, we transform the usual (x, y) coordinates to $\xi = x - y$ and $\eta = x + y$, and introduce

$$\alpha(\xi) = \begin{cases} \dfrac{2}{\pi}\left(1 - \dfrac{\xi^2}{4}\right)^{1/2} + \dfrac{\xi}{\pi} \sin^{-1}\dfrac{\xi}{2} - \dfrac{|\xi|}{2}, & -2 \leqslant \xi \leqslant 2; \\ 0, & |\xi| > 2; \end{cases} \tag{5.3}$$

then $\eta(\xi) = 2\alpha(\xi) + |\xi|$ or

$$\begin{aligned} x &= \alpha(\xi) + \xi/2 + |\xi|/2, \\ p_\Omega(x) &= \alpha(\xi) - \xi/2 + |\xi|/2. \end{aligned} \tag{5.4}$$

We now investigate the empirical distribution of the scaled row lengths, as given by the probability measure

$$\sigma_\lambda = \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j N^{-1/2}}. \tag{5.5}$$

Proposition 5.1. *Let* σ *be the probability measure given by the weak limit of the* σ_λ *as* $N \to \infty$, *so that* σ *is absolutely continuous with probability density function* q_Ω.

(i) *Then* q_Ω *is decreasing with*

$$q_\Omega(x) = -p'_\Omega(x)/2, \qquad 0 \leqslant x \leqslant 2. \tag{5.6}$$

Further, $p_\Omega(2 - x)$ *induces* $q_\Omega(x)\mathrm{d}x$ *from the uniform distribution on* $[0, 2]$, *so that*

$$\frac{x}{2} = \int_0^{p_\Omega(2-x)} q_\Omega(u)\mathrm{d}u, \qquad 0 \leqslant x \leqslant 2. \tag{5.7}$$

In particular, σ has mean $\frac{1}{2}$.

(ii) *q_Ω has algebraic singularities at 0 and 2 and is asymptotic to*

$$q_\Omega(x) \asymp \begin{cases} \dfrac{\pi}{2}\left(\dfrac{3\pi x}{2}\right)^{-1/3}, & x \to 0+, \\ \dfrac{(2-x)^{1/2}}{2\pi}, & x \to 2-, \end{cases} \tag{5.8}$$

in the sense that the ratio of left- and right-hand sides converges to one.

Proof. (i) Suppose that (h_{λ_N}) converges to p_Ω in measure as $N \to \infty$ for some sequence of $\lambda_N \vdash N$. For any bounded and continuous real function f, we have $(n/(2N^{1/2}))\int f\,\mathrm{d}\sigma_\lambda = \int_0^2 f(h_\lambda(x))\mathrm{d}x/2$, and hence

$$\frac{n}{2N^{1/2}}\int_{\Omega_N}\int_0^2 f(y)\sigma_{\lambda_N}(\mathrm{d}y)\nu_N(\mathrm{d}\lambda) = \frac{1}{2}\int_{\Omega_N}\int_0^2 f(h_{\lambda_N}(x))\mathrm{d}x\,\nu_N(\mathrm{d}\lambda). \tag{5.9}$$

Taking the limit as $N \to \infty$, we deduce from Vershik and Kerov (1977) that

$$\int_0^2 f(y)q_\Omega(y)\mathrm{d}y = \int_0^2 f(y)\sigma(\mathrm{d}y) = \frac{1}{2}\int_0^2 f(p_\Omega(x))\mathrm{d}x, \tag{5.10}$$

and, in particular,

$$\int_0^2 xq_\Omega(x)\mathrm{d}x = \lim_{N\to\infty}\int_0^\infty x\sigma_{\lambda_N}(\mathrm{d}x) = \frac{1}{2}. \tag{5.11}$$

It is evident that $p_\Omega(x)$ is convex and decreasing on $[0, 2]$ and that its graph is symmetrical about the line $y = x$; the latter fact implies that $p_\Omega(p_\Omega(y)) = y$. On setting $x = p_\Omega(y)$ in the right-hand side of (5.10) and using the inverse function theorem, we obtain (5.6) and the equivalent form (5.7). Thus the density of the limiting distribution of the row lengths is given by the derivative of the Vershik density.

(ii) In (5.4) we can make the change of variables $\xi = -2\cos\phi$ for $0 \leqslant \phi \leqslant \pi$ so that

$$x = \frac{2}{\pi}\sin\phi - \frac{2}{\pi}\phi\cos\phi = \frac{2}{\pi}\left\{\left(\frac{1}{2!} - \frac{1}{3!}\right)\phi^3 - \left(\frac{1}{4!} - \frac{1}{5!}\right)\phi^5 + \dots\right\} \tag{5.12}$$

and from (5.6) we deduce

$$q_\Omega(x) = -\frac{1}{2}\frac{\mathrm{d}p_\Omega/\mathrm{d}\phi}{\mathrm{d}x/\mathrm{d}\phi} = \frac{\pi - \phi}{2\phi}. \tag{5.13}$$

From these identities the asymptotic expansions follow. □

Remarks. (i) Some similar observations were made in Remark 1.7 of Borodin *et al*. (2000). The Wigner semicircle law likewise has a square-root singularity, so (5.8) was to be expected from the general analogy between eigenvalues of random matrices and random permutations which is considered by Baik *et al*. (1999). The density is unbounded at 0 since only a logarithmic repulsion term prevents the charges from accumulating there.

(ii) In the next section we shall consider a particular random matrix ensemble for which the asymptotic eigenvalue distribution has such an equilibrium configuration. The ensemble in question is an orthogonal ensemble, although similar techniques work for unitary and symplectic ensembles.

6. Generalized orthogonal ensemble: concentration of measure

Let $M_n^s(\mathbb{R})$ be the space of real symmetric $n \times n$ matrices and $M_n^+(\mathbb{R})$ the convex open subset of strictly positive definite matrices. Let $X^\dagger$ be the transpose of $X \in M_n(\mathbb{R})$, let τ denote the trace functional $\tau(X) = \mathrm{trace}(X)$ and $\|X\|_{M_n}$ the usual operator norm. The Hilbert–Schmidt norm is $\|X\|_{c^2} = \tau(X^\dagger X)^{1/2}$. For $X \in M_n^s(\mathbb{R})$, and υ a real function defined on the spectrum of X, we form $\upsilon(X)$ by functional calculus; in particular, we can take $\upsilon(x) = \log\Gamma(x+1)$ and introduce $V(X) = \tau(\upsilon(X))$ for $X \in M_n^+(\mathbb{R})$.

We can take $\mathrm{d}X$ to be the product of the standard Lebesgue measure on the matrix entries on or above the leading diagonal, and form the probability measure

$$\nu_n(\mathrm{d}X) = Z_n^{-1}\exp\{-V(nX)\}\mathbb{I}_{M_n^+(\mathbb{R})}(X)\mathrm{d}X \tag{6.1}$$

for some normalizing constant Z_n. By the estimates of Boutet de Monvel *et al*. (1995), there exist $0 < c$, $K < \infty$ such that $\nu_n\{X \in M_n^+(\mathbb{R}) : \|X\|_{M_n} > K\} \leqslant \mathrm{e}^{-cn}$ for all sufficiently large n; see also Corollary 4.4. We let $\nu_n^{(K)}$ be the conditional probability measure that arises from conditioning ν_n on $\{X \in M_n^+(\mathbb{R}) : \|X\|_{M_n} \leqslant K\}$ for some $K > 2$. This ensemble is the matrix analogue of Plancherel measure on the Young diagrams, for reasons which we shall now explain.

For $X \in M_n^+(\mathbb{R})$, there exists a real orthogonal matrix U such that $U^\dagger XU$ is the diagonal matrix D_λ with diagonal entries $\lambda = (\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0)$ given by the decreasing list of eigenvalues with multiplicity. The ensemble $\nu_n^{(K)}(\mathrm{d}X)$ is invariant under the conjugation action $(X \mapsto U^\dagger XU)$ of the orthogonal matrices on $M_n^+(\mathbb{R})$, hence the term 'orthogonal ensemble'. It follows by standard arguments that the eigenvalue map $\Lambda : X \mapsto D_\lambda$ induces from $\nu_n^{(K)}(\mathrm{d}X)$ the probability measure

$$\omega_n^{(K)}(\mathrm{d}\lambda) = Z_n^{-1}\exp\left\{-\sum_{j=1}^n \upsilon(n\lambda_j)\right\}\prod_{j,k:k>j}|\lambda_j - \lambda_k|\mathbb{I}_{\Delta_K^n}(\lambda)\mathrm{d}\lambda_1\mathrm{d}\lambda_2\ldots\mathrm{d}\lambda_n \tag{6.2}$$

on the simplex $\Delta_K^n = \{\lambda \in \mathbb{R}_+^n : K \geqslant \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n > 0\}$, where the factor $\prod_{j,k:k>j}(\lambda_j - \lambda_k)$ arises from the Jacobian of the transformation; see Mehta (1991). The potential resembles (4.1), or (1.1) with $\ell_j/n \leftrightarrow \lambda_j$.

This ensemble is a variant of the generalized orthogonal ensemble, as considered by Dyson (1962), Mehta and many others. While the potential $V(X)$ is scaled to $V(nX)$ in (6.1), it follows from Stirling's formula that, as $n \to \infty$,

$$n\log\Gamma(nx+1) \sim n^2 x\log nx/\mathrm{e} \sim n^2\log\Gamma(x+1) + n^2\log n, \qquad x \geqslant 1;$$

so, after normalization, the asymptotic properties of the ensemble (6.1) are similar to those of the generalized orthogonal ensemble of Boutet de Monvel *et al*. (1995, p. 601) in which the potential is scaled to $nV(X)$ on $M_n^s(\mathbb{R})$. Our main result is a concentration inequality which improves with increasing dimension. The potential $v(x) = \log \Gamma(x+1)$ has $v''(x) \to 0$ as $x \to \infty$, so we need to refine considerably the arguments of Blower (2001, Section 3).

Theorem 6.1. (i) *Let* $F : (M_n^+(\mathbb{R}), c^2) \to \mathbb{R}$ *be an L-Lipschitz function such that* $\int_{M_n^+(\mathbb{R})} F(X)\nu_n^{(K)}(\mathrm{d}X) = 0$. *Then*

$$\int_{M_n^+(\mathbb{R})} \exp\{tF(X)\}\nu_n^{(K)}(\mathrm{d}X) \leqslant \exp\{(Kn+1)L^2t^2/n^2\}, \qquad t \in \mathbb{R}. \tag{6.3}$$

(ii) *Let* $G : (\Delta_K^n, \mathbb{R}^n) \to \mathbb{R}$ *be an L-Lipschitz function such that* $\int_{\Delta_K^n} G(\lambda)\omega_n^{(K)}(\mathrm{d}\lambda) = 0$. *Then*

$$\int_{\Delta_K^n} \exp\{tG(\lambda)\}\omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \exp\{(Kn+1)L^2t^2/n^2\}, \qquad t \in \mathbb{R}. \tag{6.4}$$

The proof of Theorem 6.1 depends upon the following non-commutative extension of Lemma 4.1; the reader may compare (6.5) with (4.2). Guionnet and Zeitouni (2000) prove a version for general convex functions v. We refer the reader to Horn and Johnson (1991) for a general characterization of matrix monotone functions.

Lemma 6.2. *Let* $v(x) = \log \Gamma(1+x)$ *and* $V(X) = \tau(v(X))$. *Then:*

(i) $X \mapsto v(X)$ *is operator convex on* $M_n^+(\mathbb{R})$;
(ii) V *is a locally uniformly convex function on* $M_n^+(\mathbb{R})$, *in the sense that*

$$sV(X) + (1-s)V(Y) - V(sX + (1-s)Y) \geqslant \frac{s(1-s)}{4(K+1)}\|X - Y\|_{c^2}^2, \qquad 0 \leqslant s \leqslant 1, \tag{6.5}$$

holds whenever X *and* Y *in* $M_n^+(\mathbb{R})$ *have operator norm less than or equal to* K.

Proof of Lemma 6.2. The proof involves a non-commutative analogue of Lemma 3.3 in the form of a complicated integral formula which we derive and then simplify.

(i) It follows from Euler's product formula, as in Copson (1935), that

$$v(z) = -\log(1+z) + \sum_{k=1}^{\infty}\left\{(1+z)\log\left(1+\frac{1}{k}\right) - \log\left(1+\frac{z+1}{k}\right)\right\} \tag{6.6}$$

$$= \int_1^{\infty}\left(\frac{-1}{t} + \frac{1}{t+z}\right)\mathrm{d}t + \sum_{k=1}^{\infty}\left\{(1+z)\log\left(1+\frac{1}{k}\right) + \int_1^{\infty}\left(\frac{-1}{t} + \frac{k}{kt+1+z}\right)\mathrm{d}t\right\};$$

this involves positive combinations of z and $(t+z)^{-1}$ for $t > 1$. Since these are operator convex functions on $(0, \infty)$, we deduce that

$$sv(X) + (1-s)v(Y) - v(sX + (1-s)Y) \geqslant 0, \qquad 0 \leqslant s \leqslant 1, \tag{6.7}$$

holds for all matrices with $X, Y \geqslant 0$ in the sense of operators; hence v is operator convex.

(ii) Let us pause to observe that for $A, B \in M_n^s(\mathbb{R})$ with $A \geqslant \delta I \geqslant 0$, the trace satisfies $\tau(ABABA) \geqslant \delta^3\tau(B^2)$ by commutativity.

To establish local uniform convexity of V, we introduce

$$\psi(s) = \tau\{(tI + sX + (1-s)Y)^{-1}\}, \qquad 0 \leqslant s \leqslant 1, \tag{6.8}$$

and use the mean value theorem to obtain the identity

$$s\psi(1) + (1-s)\psi(0) - \psi(s) = \frac{s(1-s)}{2}\psi''(\bar{s})$$

for some $\bar{s}$ with $0 < \bar{s} < 1$. One can calculate the derivatives to deduce

$$\begin{aligned} &s\tau\{(tI + X)^{-1}\} + (1-s)\tau\{(tI + Y)^{-1}\} - \tau\{(tI + sX + (1-s)Y)^{-1}\} \\ &= \frac{s(1-s)}{2}\tau\{(tI + Z)^{-1}(X - Y)(tI + Z)^{-1}(X - Y)(tI + Z)^{-1}\}, \end{aligned} \tag{6.9}$$

where $Z = \bar{s}X + (1-\bar{s})Y$, and then obtain, from the preceding observation about the trace,

$$(6.9) \geqslant \frac{s(1-s)}{2(1+K)^3}\tau\{(X - Y)^2\}, \tag{6.10}$$

since $(tI + Z)^{-1} \geqslant (t + K)^{-1}I$.

Consequently, the integral formula (6.6) gives

$$\begin{aligned} &sV(X) + (1-s)V(Y) - V(sX + (1-s)Y) \\ &= \int_1^\infty (s\tau\{(tI + X)^{-1}\} + (1-s)\tau\{(tI + Y)^{-1}\} - \tau\{(tI + sX + (1-s)Y)^{-1}\})\mathrm{d}t \\ &\quad + \sum_{k=1}^\infty \int_1^\infty \Big(s\tau\{((kt+1)I + X)^{-1}\} + (1-s)\tau\{((kt+1)I + Y)^{-1}\} \\ &\quad - \tau\{((kt+1)I + sX + (1-s)Y)^{-1}\}\Big)k\,\mathrm{d}t \\ &\geqslant \frac{s(1-s)}{2}\left\{\int_1^\infty \frac{\mathrm{d}t}{(t+K)^3} + \sum_{k=1}^\infty \frac{k\,\mathrm{d}t}{(kt+1+K)^3}\right\}\|X - Y\|_{c^2}^2. \end{aligned} \tag{6.11}$$

We evaluate these integrals and obtain

$$\begin{aligned} (6.11) &\geqslant \frac{s(1-s)}{2}\left\{\frac{1}{2(1+K)^2} + \sum_{k=1}^\infty \frac{1}{2(k+1+K)^2}\right\}\|X - Y\|_{c^2}^2 \\ &\geqslant \frac{s(1-s)}{4(1+K)}\|X - Y\|_{c^2}^2. \end{aligned} \tag{6.12}$$

□

Proof of Theorem 6.1. (i) The transportation inequality

$$\mathrm{Tc}_2(\mu, \nu_n^{(K)}) \leqslant \frac{4(nK+1)}{n^2} \operatorname{Ent}(\mu|\nu_n^{(K)}) \tag{6.13}$$

with cost function $\|\cdot\|_{c^2}^2$ follows from Lemma 6.2(ii) as in Theorem 3.2, and we gain the advantageous constant $4(Kn+1)/n^2$ in this inequality by scaling X and Y in (6.5). This transportation inequality implies the concentration inequality (6.3) by the following argument; similar techniques have been used by Bobkov and Götze (1999).

The required concentration inequality

$$\int_{M_n^+(\mathbb{R})} \exp\left\{ tF(X) - \frac{(Kn+1)L^2 t^2}{n^2} \right\} \nu_n^{(K)}(\mathrm{d}X) \leqslant 1 \tag{6.14}$$

holds for all L-Lipschitz functions F with $\int_{M_n^+(\mathbb{R})} F(X)\nu_n^{(K)}(\mathrm{d}X) = 0$, if and only if

$$\operatorname{Ent}(\rho|\nu_n^{(K)}) \geqslant t\int_{M_n^+(\mathbb{R})} F(X)\rho(\mathrm{d}X) - \frac{(Kn+1)L^2 t^2}{n^2} \qquad (t > 0) \tag{6.15}$$

holds for all such F and all probability measures ρ that are of finite relative entropy with respect to $\nu_n^{(K)}$. Optimizing over t, we see that the preceding inequality is equivalent to the inequality

$$\frac{4(Kn+1)}{n^2} \operatorname{Ent}(\rho|\nu_n^{(K)}) \geqslant \frac{1}{L^2} \left(\int_{M_n^+(\mathbb{R})} F(X)\rho(\mathrm{d}X) - \int_{M_n^+(\mathbb{R})} F(X)\nu_n^{(K)}(\mathrm{d}X) \right)^2 \tag{6.16}$$

holding for all L-Lipschitz functions. By the Kantorovich–Rubinstein duality formula, this is equivalent to the transportation inequality

$$\frac{4(Kn+1)}{n^2} \operatorname{Ent}(\rho|\nu_n^{(K)}) \geqslant \mathrm{Tc}_1(\rho, \nu_n^{(K)})^2$$

for the cost function $\|\cdot\|_{c^2}$. This follows from the quadratic transportation inequality (6.13) by the Cauchy–Schwarz inequality.

(ii) The eigenvalue map $\Lambda : (M_n^+(\mathbb{R}), c^2) \to (\Delta^n, \mathbb{R}^n)$ induces $\omega_n^{(K)}$ from $\nu_n^{(K)}$ and is 1-Lipschitz by Lidskii's theorem; see Simon (1979). Hence we can deduce (6.4) from (6.3) by taking $F(X) = G(\Lambda(X))$ and following the proof from Blower (2001, Section 4). Alternatively, one can argue as in the proofs of Proposition 4.3 and Corollary 3.3. □

Corollary 6.3. *Let* $\bar{\lambda}_j = \int_{\Delta_K^n} \lambda_j \omega_n^{(K)}(\mathrm{d}\lambda)$ *be the mean of the jth largest eigenvalue.*

(i) *Then*

$$\int_{\Delta_K^n} \exp\left\{ t^2 \sum_{j=1}^n (\lambda_j - \bar{\lambda}_j)^2 \right\} \omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \exp\{4t^2(K + n^{-1})\}, \qquad t^2 \leqslant n^2/(4(nK+1)). \tag{6.17}$$

(ii) *Moreover, the largest eigenvalue satisfies*

$$\int_{\Delta_K^n} \exp\{t^2 n(\lambda_1 - \bar{\lambda}_1)^2\} \omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \{1 - 4t^2(K + n^{-1})\}^{1/2}, \qquad t^2 \leqslant n/(4(nK+1)). \tag{6.18}$$

Proof. (i) Let $\gamma(\mathrm{d}\xi) = (2\pi)^{-1/2}\mathrm{e}^{-\xi^2/2}\mathrm{d}\xi$ be the $N(0, 1)$ Gaussian distribution and $\gamma^{\otimes n}$ be its n-fold product on $\mathbb{R}^n$. Then we can write

$$\int_{\Delta_K^n} \exp\left\{ t^2 \sum_{j=1}^n (\lambda_j - \bar{\lambda}_j)^2/2 \right\} \omega_n^{(K)}(\mathrm{d}\lambda)$$

$$= \int_{\Delta_K^n} \int_{\mathbb{R}^n} \exp\left\{ t \sum_{j=1}^n \xi_j(\lambda_j - \bar{\lambda}_j) \right\} \gamma^{\otimes n}(\mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n)\omega_n^{(K)}(\mathrm{d}\lambda). \tag{6.19}$$

The next step is to interchange the order of integration and bound the resulting integrand by

$$\int_{\Delta_K^n} \exp\left\{ t \sum_{j=1}^n \xi_j(\lambda_j - \bar{\lambda}_j) \right\} \omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \exp\left\{ \frac{(nK+1)t^2}{n^2} \sum_{j=1}^n \xi_j^2 \right\}, \tag{6.20}$$

which follows from (6.4). Hence we have

$$\int_{\Delta_K^n} \exp\left\{ \frac{t^2}{2} \sum_{j=1}^n (\lambda_j - \bar{\lambda}_j)^2 \right\} \omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \int_{\mathbb{R}^n} \exp\left\{ \frac{(nK+1)t^2}{n^2} \sum_{j=1}^n \xi_j^2 \right\} \gamma^{\otimes n}(\mathrm{d}\xi_1 \ldots \mathrm{d}\xi_n)$$

$$= \left(1 - \frac{2t^2(nK+1)}{n^2} \right)^{-n/2}. \tag{6.21}$$

By elementary estimates on $n\log(1 - x/n)$ we can deduce the stated result (6.17).

(ii) Inequality (6.18) is proved by a similar but easier argument. □

Remark. The main virtue of inequality (6.17) is the lack of any scaling on the sum on the left-hand side, while the exponent on the right-hand side improves as n increases.

7. Mean field theory and weak convergence

Let $\sigma_n^{(\lambda)} = (1/n)\sum_{j=1}^n \delta_{\lambda_j}$ be the empirical eigenvalue distribution, where typically the λ_j are random subject to $\omega_n^{(K)}$. We let $\bar{\sigma}_n$ be the integrated density of states; that is, the probability measure defined via F. Riesz's theorem by

$$\int_{[0,K]} f(x)\bar{\sigma}_n(\mathrm{d}x) = \int_{\Delta_K^n} \int_{[0,K]} f(x)\sigma_n^{(\lambda)}(\mathrm{d}x)\omega_n^{(K)}(\mathrm{d}\lambda) \tag{7.1}$$

for all continuous $f : [0, K] \to \mathbb{R}$. See Boutet de Monvel *et al*. (1995) for a general discussion.

Proposition 7.1. *Under the laws $\omega_n^{(K)}$, the empirical distributions of eigenvalues converge weakly almost surely in the sense that*

$$\int_0^K f(x)\sigma_n^{(\lambda)}(\mathrm{d}x) - \int_0^K f(x)\bar{\sigma}_n(\mathrm{d}x) \to 0, \qquad \text{as } n \to \infty, \tag{7.2}$$

almost surely with respect to $\otimes_{n=1}^\infty \omega_n^{(K)}$ for all continuous real functions f.

Proof. By the Stone–Weierstrass theorem, it suffices to prove this for an L-Lipschitz function f. Then

$$G_n(\lambda) = \int f\, \mathrm{d}\sigma_n^{(\lambda)} - \int f\, \mathrm{d}\overline{\sigma}_n$$

is an $(Ln^{-1/2})$-Lipschitz function $\Delta_K^n \to \mathbb{R}$ by the Cauchy–Schwarz inequality; further, $\int G_n(\lambda)\omega_n^{(K)}(\mathrm{d}\lambda) = 0$ holds by the definition of $\overline{\sigma}_n$.

We deduce from Theorem 6.1(ii) that, for each $\varepsilon > 0$,

$$\omega_n^{(K)}\{\lambda \in \Delta_K^n : |G_n(\lambda)| \geqslant \varepsilon\} \leqslant \exp\{-\varepsilon^2 n^3/(4CL^2(Kn+1))\}, \qquad n \geqslant 1. \tag{7.3}$$

Since the right-hand side is of rapid decay as $n \to \infty$, it follows from the first Borel–Cantelli lemma that

$$\left\{(\lambda_n) \in \prod_{n=1}^{\infty} \Delta_K^n : |G_n(\lambda_n)| \geqslant \varepsilon \text{ for infinitely many } n\right\}$$

has measure zero with respect to $\otimes_{n=1}^{\infty}\omega_n^{(K)}$. Hence $G_n(\lambda_n) \to 0$ almost surely as $n \to \infty$. □

According to the mean-field theory of Johansson (1998) and Boutet de Monvel *et al.* (1995), the $\overline{\sigma}_n$ are also weakly convergent as $n \to \infty$, and their limit is the measure $q_\Omega(x)\mathrm{d}x$ of Section 5. For the sake of completeness, we reproduce some of this material in the style of transportation theory before presenting new applications. The potential for the eigenvalue distribution (6.2) is, for some constant C_n,

$$V(n\lambda) = \sum_{j=1}^{n} \log\Gamma(n\lambda_j + 1) \sum_{1 \leqslant j < k \leqslant n} \log(\lambda_j - \lambda_k) + C_n. \tag{7.4}$$

This we approximate by the mean-field Hamiltonian

$$H_n(\lambda) = \sum_{j=1}^{n}\left(\log\Gamma(n\lambda_j + 1) - n\int_0^K \log|\lambda_j - y| q_n(y)\mathrm{d}y\right), \tag{7.5}$$

where q_n is a probability density function on $[0, K]$ that satisfies

$$q_n(x) = \mathrm{Z}_n^{-1}\exp\left(-v_n(x) + n\int_0^K \log|x - y| q_n(y)\mathrm{d}y\right), \qquad x \in [0, K], \tag{7.6}$$

with $K > 2$ and $v_n(x) = \log\Gamma(nx+1) + s_n x$; here $s_n = nc_1$, where c_1 is a Lagrange multiplier. Using convexity in a suitable Hilbert space, Boutet de Monvel *et al.* (1995) prove the existence of q_n as the unique minimizer of

$$\Psi_n(q) = \frac{1}{2}\iint_{[0,K]^2} \log\frac{1}{|x-y|}\, q(x)q(y)\mathrm{d}x\,\mathrm{d}y$$
$$+\frac{1}{n}\log\left\{\int_0^K \exp\left(-v_n(x) + n\int_0^K \log|x-y|q(y)\mathrm{d}y\right)\mathrm{d}x/\mathrm{K}\right\} \tag{7.7}$$

amongst probability measures on $[0, K]$ of finite logarithmic energy. The q_n converge pointwise as $n \to \infty$ to a probability density function q_Ω that satisfies

$$\int_0^K \log|\lambda - y| q_\Omega(y)\mathrm{d}y = \lambda\log\lambda/\mathrm{e} + c_1\lambda - c_2, \tag{7.8}$$

for constants c_1 and c_2, on the support of q_Ω. Since $\lambda\log\lambda/\mathrm{e} + c_1\lambda - c_2$ is convex on $(0, \infty)$, the support consists of a single interval as in Boutet de Monvel *et al*. (1995, p. 592). We can identify the constant c_2 by integrating to obtain

$$c_2 = \int_0^K (x\log x/\mathrm{e} + c_1 x) q_\Omega(x)\mathrm{d}x - \iint_{[0,K]^2} \log|x-y| q_\Omega(x)q_\Omega(y)\mathrm{d}x\,\mathrm{d}y. \tag{7.9}$$

The constant c_1 may be adjusted so as to change the value of the mean $\int xq_\Omega(x)\mathrm{d}x$. We can regard q_Ω as a minimizer of the variational problem

$$\inf_q \left\{\frac{1}{2}\iint_{(0,\infty)^2} \log\frac{1}{|x-y|}\, q(x)q(y)\mathrm{d}x\,\mathrm{d}y + \int_0^\infty x\log\frac{x}{\mathrm{e}}\, q(x)\mathrm{d}x\right\}$$

over all probability density functions q on $(0, \infty)$ that have finite logarithmic energy and that are subject to the constraint $\int_0^\infty xq(x)\mathrm{d}x = c_3$. On account of Proposition 5.1, we take $c_3 = \frac{1}{2}$. (The similar variational problem considered in Logan and Shepp (1997, p. 213) involves a different range of integration in the quadratic functional and a different function space.)

The logarithmic interaction of the eigenvalues is so feeble that they decouple in the limit as $n \to \infty$, in the sense of the following theorem, which expresses quantitatively the *weak factorization* of the integrated density of states.

Theorem 7.2. *There exists a sequence of product probability measures $\tilde{\omega}_n$ on Δ_K^n such that*

$$\int_{\Delta_K^n} F_n(\lambda)\omega_n^{(K)}(\mathrm{d}\lambda) - \int_{\Delta_K^n} F_n(\lambda)\tilde{\omega}_n(\mathrm{d}\lambda) \to 0, \qquad \textit{as } n \to \infty, \tag{7.10}$$

for all sequences of $(Ln^{-1/2})$-Lipschitz functions $F_n : \Delta_K^n \to \mathbb{R}$ and all L with $0 < L < \infty$.

Proof. We introduce the Hamiltonian

$$\tilde{H}_n(\lambda) = \sum_{j=1}^n \left(v_n(\lambda_j) - n\int_0^K \log|\lambda_j - y| q_\Omega(y)\mathrm{d}y\right) \tag{7.11}$$

which differs from $H_n(\lambda)$ only in that it involves the equilibrium configuration q_Ω instead of q_n and includes the Lagrange multiplier. The corresponding Gibbs probability measure is $\tilde{\omega}_n(\mathrm{d}\lambda) = \tilde{Z}_n^{-1}\mathrm{e}^{-\tilde{H}_n(\lambda)}\,\mathrm{d}\lambda$ on Δ_K^n, where by symmetry

$$\tilde{Z}_n = \frac{1}{n!}\left\{\int_0^K \exp\left(-v_n(x) + n\int_0^K \log|x-y|q_\Omega(y)\mathrm{d}y\right)\mathrm{d}x/K\right\}^n. \tag{7.12}$$

By Stirling's formula and (7.8), we can write the exponent in (7.12) as

$$\begin{aligned}&-v_n(x) + n\int_0^K \log|x-y|q_\Omega(y)\mathrm{d}y\\ &\quad = -nc_2 - \tfrac{1}{2}\log 2\pi n - \tfrac{1}{2}\log x - nx\log n + O((1+nx)^{-1}),\end{aligned}$$

as $n\to\infty$, and so we can estimate the integral in (7.12) by using

$$\int_0^K \frac{\exp(-nx\log n)}{x^{1/2}}\mathrm{d}x = \left(\frac{\pi}{n\log n}\right)^{1/2} + O\left(\frac{\exp(-Kn\log \mathrm{n})}{K^{1/2}n\log n}\right), \qquad n\to\infty.$$

Hence, we have

$$\log\tilde{Z}_n = -n^2c_2 + O(n\log n) \tag{7.13}$$

as $n\to\infty$, where c_2 is given by (7.9).

Let $F_n : \Delta_K^n \to \mathbb{R}$ be an $(Ln^{-1/2})$-Lipschitz function. Then the difference between the means of F_n with respect to the true and approximate Gibbs measures satisfies

$$\begin{aligned}\left|\int_{\Delta_K^n} F_n(\lambda)\omega_n^{(K)}(\mathrm{d}\lambda) - \int_{\Delta_K^n} F_n(\lambda)\tilde{\omega}_n(\mathrm{d}\lambda)\right| &\leqslant Ln^{-1/2}\mathrm{Tc}_1(\tilde{\omega}_n, \omega_n^{(K)})\\ &\leqslant Ln^{-1/2}\{\mathrm{Tc}_2(\tilde{\omega}_n, \omega_n^{(K)})\}^{1/2}\end{aligned} \tag{7.14}$$

by the Kantorovich–Rubinstein duality formula (2.3) and the Cauchy–Schwarz inequality. Further, by the transportation inequality (6.6) for the Gibbs measure,

$$(7.14) \leqslant \frac{L}{n^{1/2}}\left\{\frac{4(nK+1)}{n^2}\mathrm{Ent}(\tilde{\omega}_n|\omega_n^{(K)})\right\}^{1/2}. \tag{7.15}$$

Now $\tilde{H}_n(\lambda)$ and $V(n\lambda)$ involve the same $\log\Gamma$ terms, so the relative entropy satisfies

$$\begin{aligned}\mathrm{Ent}(\tilde{\omega}_n|\omega_n^{(K)}) &= \int_{\Delta_K^n}(\log \mathrm{Z}_n - \log\tilde{Z}_n + V(n\lambda) - \tilde{H}_n(\lambda))\tilde{\omega}_n(\mathrm{d}\lambda)\\ &= \log \mathrm{Z}_n - \log\tilde{Z}_n - \int_{\Delta_K^n}\sum_{j,k:j<k}\log|\lambda_j-\lambda_k|\tilde{\omega}_n(\mathrm{d}\lambda)\\ &\quad + \int_{\Delta_K^n}\int_0^K\sum_{j=1}^n n\log|\lambda_j - y|q_\Omega(y)\mathrm{d}y\,\tilde{\omega}_n(\mathrm{d}\lambda).\end{aligned} \tag{7.16}$$

To control the normalizing constant Z_n, we introduce the probability measure $Q_n(\mathrm{d}\lambda) = n!q_\Omega(\lambda_1)\dots q_\Omega(\lambda_n)\mathrm{d}\lambda_1\dots\mathrm{d}\lambda_n$ on Δ_K^n and apply Jensen's inequality to obtain

$$0 \leqslant \mathrm{Ent}(\omega_n^K | Q_n) = \int_{\Delta_K^n} \log \frac{\mathrm{d}\omega_n^{(K)}}{\mathrm{d}Q_n} \omega_n^{(K)}(\mathrm{d}\lambda) \tag{7.17}$$

$$= -\log n! - \log Z_n - \int_{\Delta_K^n} \left\{ V(n\lambda) + \sum_{j=1}^n \log q_\Omega(\lambda_j) \right\} \omega_n^{(K)}(\mathrm{d}\lambda).$$

The final sum here is $O(n)$, since Proposition 7.1 gives

$$\int_{\Delta_K^n} \frac{1}{n} \sum_{j=1}^n \log q_\Omega(\lambda_j) \omega_n^{(K)}(\mathrm{d}\lambda) \to \int q_\Omega(x) \log q_\Omega(x) \mathrm{d}x, \qquad \text{as } n \to \infty,$$

where the latter integral is finite by Proposition 5.1. We have, on substituting (7.13) and (7.17) into (7.16),

$$\mathrm{Ent}(\tilde{\omega}_n | \omega_n^{(K)}) \tag{7.18}$$

$$\leqslant n^2 \int_0^K (x \log x/\mathrm{e} + c_1 x) q_\Omega(x) \mathrm{d}x - n^2 \int_0^K \int_0^K \log|x-y| q_\Omega(x) q_\Omega(y) \mathrm{d}x \, \mathrm{d}y$$

$$- n \int_{\Delta_K^n} \left(\sum_{j=1}^n \lambda_j \log \lambda_j/\mathrm{e} + c_1 \lambda_j \right) \omega_n^{(K)}(\mathrm{d}\lambda) - \int_{\Delta_K^n} \sum_{j,k\,:j<k} \log|\lambda_j - \lambda_k| \tilde{\omega}_n(\mathrm{d}\lambda)$$

$$+ \int_{\Delta_K^n} \int_0^K \sum_{j=1}^n n \log|\lambda_j - y| q_\Omega(y) \mathrm{d}y \, \tilde{\omega}_n(\mathrm{d}\lambda) + \int_{\Delta_K^n} \sum_{j,k:j<k} \log|\lambda_j - \lambda_k| \omega_n^{(K)}(\mathrm{d}\lambda)$$

$$+ \mathrm{O}(n \log n).$$

First we consider the latest integral. As in classical potential theory, we introduce

$$\delta_n = \sup_\lambda \left\{ \binom{n}{2}^{-1} \sum_{j,k:1 \leqslant j<k \leqslant n} \log(\lambda_j - \lambda_k) - \frac{1}{n} \sum_{j=1}^n \int_0^K \log|\lambda_j - y| q_\Omega(y) \mathrm{d}y : \lambda \in \Delta_K^n \right\} \tag{7.19}$$

and define the *Fekete points* at stage n to be points (λ_j) at which the supremum is attained. Further, we observe that the sequence $(\delta_n)_{n=1}^\infty$ is decreasing; this follows from a simple counting argument as in Ransford (1995, p. 153). The atomic probability measures, that assign mass $1/n$ to each of the Fekete points at stage n, converge weakly to the equilibrium distribution $q_\Omega(x)\mathrm{d}x$ as $n \to \infty$; see Johansson (1998). As in (5.7), one can show that q_Ω is a continuous function on (0, K] and that there exists $\alpha < 1$ such that $x^\alpha q_\Omega(x) \to 0$ as $x \to 0+$. Hence $\int \log|\lambda - y| q_\Omega(y) \mathrm{d}y$ is a continuous and subharmonic function of λ; so one can follow the proof from Ransford (1995) to show that $\delta_n \to 0$ as $n \to \infty$.

Boutet de Monvel *et al.* (1995, p. 606 (ii)) show that $\int \lambda \overline{\sigma}_n(\mathrm{d}\lambda) \to \int \lambda q_\Omega(\lambda) \mathrm{d}\lambda$ and $\int \lambda \log(\lambda/\mathrm{e}) \overline{\sigma}_n(\mathrm{d}\lambda) \to \int \lambda \log(\lambda/\mathrm{e}) q_\Omega(\lambda) \mathrm{d}\lambda$ as $n \to \infty$. Further, they show convergence of the logarithmic energy

$$\iint_{[0,K]^2} \log|s-t| q_n(s) q_n(t) \mathrm{d}s\, \mathrm{d}t \to \iint_{[0,K]^2} \log|s-t| q_\Omega(s) q_\Omega(t) \mathrm{d}s\, \mathrm{d}t$$

as $n \to \infty$.

Hence the sum in (7.18) is $o(n^2)$ as $n \to \infty$. When combined with (7.15), this implies the limit (7.10). □

For some purposes, it is convenient to relax the ordering assumptions on the eigenvalues. In terms of the Gibbs measure (6.2), we enlarge the phase space to $[0, K]^n$ and renormalize by dividing by $n!$. As the potential is symmetrical with respect to permutation of the $\lambda_1, \dots, \lambda_n$, we shall continue to use the notation $\omega_n^{(K)}(\mathrm{d}\lambda)$.

Corollary 7.3. *Let $\rho_{n,k}$ be the k-point correlation function of the ensemble $\omega_n^{(K)}$, that is, $n!/(n-k)!$ times the probability density function of $\lambda_1, \dots, \lambda_k$, where $\lambda_1, \dots, \lambda_n$ are the unordered random eigenvalues subject to $\omega_n^{(K)}$ on $[0, K]^n$. Then*

$$\frac{(n-k)!}{n!} \int_{[0,K]^n} f(\lambda_1, \dots, \lambda_k) \rho_{n,k}(\lambda_1, \dots, \lambda_k) \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_k$$
$$\to \int_{[0,K]^k} f(\lambda_1, \dots, \lambda_k) q_\Omega(\lambda_1) \dots q_\Omega(\lambda_k) \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_k \quad (7.20)$$

as $n \to \infty$ for all continuous functions $f : [0, K]^k \to \mathbb{R}$.

Weak convergence of the integrated density of states is the special case $k = 1$, which involves linear statistics.

Proof. It suffices to prove this for an L-Lipschitz function f; further, we may assume that $f(\lambda_1, \dots, \lambda_k)$ is symmetrical with respect to permutation of the variables $\lambda_1, \dots, \lambda_k$. For each vector $\lambda = (\lambda_1, \dots, \lambda_n) \in [0, K]^n$ and each subset α of $[n]$ that has k elements, we introduce the (partial) vector $\lambda^\alpha = (\lambda_{\alpha_1}, \dots, \lambda_{\alpha_k})$. The real function

$$F_n(\lambda) = \frac{(n-k)!k!}{n!} \sum_{\alpha:\alpha\subseteq[n],\#\alpha=k} f(\lambda^\alpha), \qquad \lambda \in [0, K]^n, \quad (7.21)$$

is Lipschitz with constant $kLn^{-1/2}$, since only about k/n of the sets α involve any given index $j \in [n]$.

By definition of the k-point function, we have

$$\int_{[0,K]^n} F_n(\lambda) \omega_n^{(K)}(\mathrm{d}\lambda) = \frac{(n-k)!}{n!} \int_{[0,K]^k} f(\lambda_1, \dots, \lambda_k) \rho_{n,k}(\lambda_1, \dots, \lambda_k) \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_k. \quad (7.22)$$

By Theorem 7.2 and equation (7.6), we can replace the measure in the left-hand side by the product measure $q_n(\lambda_1) \dots q_n(\lambda_n) \mathrm{d}\lambda_1 \dots \mathrm{d}\lambda_n$ and only change the expression by $o(1)$ as $n \to \infty$. Since $q_n(\lambda)\mathrm{d}\lambda$ converges weakly to $q_\Omega(\lambda)\mathrm{d}\lambda$ as $n \to \infty$, the required result follows. □

Corollary 7.4. *For any continuous function* $f : [0, K]^k \to \mathbb{R}$ *that is symmetrical with respect to permutation of variables, the functions* F_n *of* (7.21) *satisfy*

$$\frac{1}{n}\log\int_{[0,K]^n}\exp\{nF_n(\lambda)\}\omega_n^{(K)}(\mathrm{d}\lambda) \to \int_{[0,K]^k} f(\lambda_1, \ldots, \lambda_k)q_\Omega(\lambda_1)\ldots q_\Omega(\lambda_k)\mathrm{d}\lambda_1\ldots\mathrm{d}\lambda_k \tag{7.23}$$

as $n \to \infty$.

Proof. It suffices to prove this for L-Lipschitz functions f. As in the preceding proof, F_n is then $(kLn^{-1/2})$-Lipschitz on $[0, K]^n$ and hence by Theorem 6.1(ii) satisfies

$$\int_{[0,K]^n}\exp\left\{nF_n(\lambda) - n\int_{[0,K]^n}F_n(\xi)\omega_n^{(K)}(\mathrm{d}\xi)\right\}\omega_n^{(K)}(\mathrm{d}\lambda) \leqslant \exp\left\{\frac{k^2(Kn+1)L^2}{n}\right\}. \tag{7.24}$$

Hence we have, for large n,

$$\frac{1}{n}\log\int_{[0,K]^n}\exp\{nF_n(\lambda)\}\omega_n^{(K)}(\mathrm{d}\lambda) = \int_{[0,K]^n}F_n(\xi)\omega_n^{(K)}(\mathrm{d}\xi) + O(1/n),$$

which implies the result by (7.20). $\square$

Proposition 7.5. *Let* X *be a random matrix subject to the Gibbs measure* $\nu_n^{(K)}$.

(i) *Then the logarithm of its mean characteristic function satisfies*

$$\frac{1}{n}\log\int_{M_n^+(\mathbb{R})}\det(zI - X)\nu_n^{(K)}(\mathrm{d}X) \to \int_0^K \log(z-x)q_\Omega(x)\mathrm{d}x \tag{7.25}$$

as $n \to \infty$, *uniformly on compact subsets of* $E = \{z \in \mathbb{C}\backslash[0, \infty) : \mathrm{dist}(z, [0, K]) > 6K^{1/2}\}$. *Here the logarithm on the right-hand side is determined by the choice* $0 < \arg z < 2\pi$, *and the choice on the left-hand side has* $\arg z^n = n\arg z$.

(ii) *Let* γ *be a simple contour that winds round* $[0, K]$, *but remains at distance greater than* $6K^{1/2}$ *from* $[0, K]$. *Then* $\chi_n(z) = \int_{M_n^+(\mathbb{R})}\det(zI - X)\nu_n^{(K)}(\mathrm{d}X)$ *has all its zeros inside* γ *for all sufficiently large* n.

Proof. (i) Let g be the complex-valued function $g(x) = \log(z - x)$ for fixed $z \in \mathbb{C}\backslash[0, \infty)$. Then g is continuously differentiable on $[0, K]$ with $|g'(x)| \leqslant L = \mathrm{dist}(z, [0, K])^{-1}$, so $L^2 < 1/(36K)$ for $z \in E$. We let $G_n(x) = \sum_{j=1}^n g(x_j)$ for $x = (x_j) \in [0, K]^n$, and we have real and imaginary parts

$$A_n(x) + \mathrm{i}B_n(x) = G_n(x) - \int_{[0,K]^n}G_n(y)\omega_n^{(K)}(\mathrm{d}y),$$

where A_n and B_n have Lipschitz constant $Ln^{1/2}$. The inequality $|\mathrm{e}^{A_n+\mathrm{i}B_n} - 1| \leqslant \mathrm{e}^{A_n}|\mathrm{e}^{\mathrm{i}B_n} - 1| + |\mathrm{e}^{A_n} - 1|$ holds by basic facts about complex numbers, and hence the estimate

$$\left|\int_{[0,K]^n} \mathrm{e}^{A_n+\mathrm{i}B_n}\omega_n^{(K)}(\mathrm{d}x)-1\right|$$
$$\leqslant \left(\int_{[0,K]^n}\mathrm{e}^{2A_n(x)}\omega_n^{(K)}(\mathrm{d}x)\right)^{1/2}\left(\int_{[0,K]^n}4\sin^2(B_n(x)/2)\omega_n^{(K)}(\mathrm{d}x)\right)^{1/2}$$
$$+\left(\int_{[0,K]^n}\mathrm{e}^{2A_n(x)}\omega_n^{(K)}(\mathrm{d}x)-1\right)^{1/2} \tag{7.26}$$

follows by the Cauchy–Schwarz inequality. By Theorem 6.1(ii) the final integral satisfies

$$0\leqslant \int_{[0,K]^n}\mathrm{e}^{2\mathrm{A}_n(x)}\omega_n^{(K)}(\mathrm{d}x)-1\leqslant \mathrm{e}^{4L^2(Kn+1)/n}-1\leqslant \mathrm{e}^{(Kn+1)/9Kn}-1<\tfrac{1}{4}, \tag{7.27}$$

for all large n; similar estimates hold for the other integrals in (7.26). Hence

$$\int_{[0,K]^n}\exp\{G_n(x)\}\omega_n^{(K)}(\mathrm{d}x)=(1+\zeta_n)\exp\left(\int_{[0,K]^n}G_n(x)\omega_n^{(K)}(\mathrm{d}x)\right) \tag{7.28}$$

where $|\zeta_n|<1$, and, by Corollary 7.3,

$$\frac{1}{n}\int_{[0,K]^n}G_n(x)\omega_n^{(K)}(\mathrm{d}x)\to\int_0^K\log(z-x)q_\Omega(x)\mathrm{d}x,\qquad \text{as } n\to\infty. \tag{7.29}$$

Since both factors on the right-hand side of (7.28) are non-zero, we can now take logarithms with appropriate choices of the argument and deduce that

$$\frac{1}{n}\log\int_{[0,K]^n}\exp\{G_n(x)\}\omega_n^{(K)}(\mathrm{d}x)\to\int_0^K\log(z-x)q_\Omega(x)\mathrm{d}x,\qquad \text{as } n\to\infty. \tag{7.30}$$

As the measure $\omega_n^{(K)}$ is induced from $\nu_n^{(K)}$ by the eigenvalue map, this gives the required result.

(ii) By the argument principle, the number of zeros of $\chi_n(z)$ that lie inside γ equals

$$\frac{1}{2\pi\mathrm{i}}\int_\gamma\frac{\chi_n'(z)}{\chi_n(z)}\,\mathrm{d}z=\frac{1}{2\pi\mathrm{i}}\int_\gamma\frac{\zeta_n'(z)}{\zeta_n(z)+1}\,\mathrm{d}z+\int_{[0,K]^n}\left\{\frac{1}{2\pi\mathrm{i}}\int_\gamma\sum_{j=1}^n\frac{\mathrm{d}z}{z-\lambda_j}\right\}\omega_n^{(K)}(\mathrm{d}\lambda), \tag{7.31}$$

where $|\zeta_n|<1$ on γ by (i). Hence both sides of (7.31) equal n. □

Remarks. (i) I do not know whether uniform convergence occurs in (7.25) for all $z\in\mathbb{C}\backslash[0,\infty)$. Unlike in the case of the unitary ensemble, where the theory of orthogonal polynomials applies, one does not know in advance that the left-hand side of (7.25) is holomorphic on $\mathbb{C}\backslash[0,\infty)$. Let us consider the simple random matrix model where $Y=I\in M_2(\mathbb{R})$ with probability $\frac{1}{2}$ and $Y=0$ with probability $\frac{1}{2}$. The mean characteristic polynomial equals $(\lambda-\frac{1}{2})^2+\frac{1}{4}$, which has complex zeros. This strange possibility seems hard to eliminate from the context of Proposition 7.5.

(ii) Tracy and Widom (1998) have obtained a determinant formula involving quaternions for the k-point correlation function of (7.20). For the determinants of random matrices with respect to the Gaussian orthogonal ensemble, Delannay and Le Caër (2000) have computed the Mellin transform of the probability density function. The formulae thus obtained do not offer a straightforward path to concentration inequalities.

Acknowledgements

I am grateful to Profs M. Ledoux, A. Guionnet and G. Roberts for helpful conversations. This work was completed while I was visiting the University of New South Wales.

References

Baik, J., Deift, P. and Johansson, K. (1999) On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, **12**, 1119–1178.

Blower, G. (2001) Almost sure weak convergence for the generalized orthogonal ensemble. *J. Statist. Phys.*, **105**, 309–335.

Bobkov, S.G. and Götze, F. (1999) Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, **163**, 1–28.

Bobkov, S.G. and Ledoux, M. (2000) From Brunn–Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, **10**, 1028–1052.

Bobkov, S.G., Gentil, I. and Ledoux, M. (2001) Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl. (9)*, **80**, 669–696.

Borodin, A., Okounkov, A. and Olshanski, G. (2000) Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.*, **13**, 481–515.

Boutet de Monvel, A., Pastur, L. and Shcherbina, M. (1995) On the statistical mechanics approach to random matrix theory: integrated density of states. *J. Statist. Phys.*, **79**, 585–611.

Copson, E.T. (1935) *An Introduction to the Theory of Functions of a Complex Variable*. Oxford: Clarendon.

Delannay, R. and Le Caër, G. (2000) Distribution of the determinant of a random real symmetric matrix from the Gaussian orthogonal ensemble. *Phys. Rev. E (3)*, **62**, 1526–1536.

Deuschel, J.-D. and Zeitouni, O. (1999) On increasing subsequences of i.i.d. samples. *Combin. Probab. Comput.*, **8**, 247–263.

Dyson, F.J. (1962) Statistical theory of the energy levels of complex systems I. *J. Math. Phys.*, **3**, 140–156.

Fulton, W.F. (1997) *Young Tableaux*, London Math. Soc. Stud. Texts 35. Cambridge: Cambridge University Press.

Gangbo, W. and McCann, R.J. (1996) The geometry of optimal transportation. *Acta Math.*, **177**, 113–161.

Guionnet, A. and Zeitouni, O. (2000) Concentration of spectral measure for large matrices. *Electron. Comm. Probab.*, **5**, 119–136.

Hardy, G.H., Littlewood, J.E. and Pólya, G. (1952) *Inequalities*, 2nd edition. Cambridge: Cambridge University Press.

Horn, R.A. and Johnson, C.R. (1991) *Matrix Analysis*. Cambridge: Cambridge Unversity Press.

Johansson, K.J. (1998) On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, **91**, 151–204.

Logan, B.F. and Shepp, L.A. (1977) A variational problem for random Young tableau. *Adv. Math.*, **26**, 206–222.

Mehta, M.L. (1991) *Random Matrices*. Second edition. San Diego: Academic Press.

Otto, F. and Villani, C. (2000) Generalization of an inequality of Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, **173**, 361–400.

Pisier, G. (1989) *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge: Cambridge University Press.

Ransford, T.J. (1995) *Potential Theory in the Complex Plane*. London Math. Soc. Stud. Texts 28. Cambridge: Cambridge University Press.

Simon, B. (1979) *Trace Ideals and Their Applications*. Cambridge: Cambridge University Press.

Talagrand, M. (1996) Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, **6**, 587–600.

Tracy, C.A. and Widom, H. (1998) Correlation functions, cluster functions, and spacing distributions for random matrices. *J. Statist. Phys.*, **92**, 809–835.

Vershik, A.M. and Kerov, S.V. (1977) Asymptotic behaviour of the Plancherel measure of the symmetric group and the limit form of Young tableau. *Dokl. Akad. Nauk. SSSR*, **233**, 1024–1027.

Villani, C. (2003) *Topics in Optimal Transportation*. Providence, RI: American Mathematical Society.

Received August 2003 and revised February 2004