On the linear statistics of Hermitian random matrices

1998 J. Phys. A: Math. Gen. 31 1141

(http://iopscience.iop.org/0305-4470/31/4/005)

J. Phys. A: Math. Gen. **31** (1998) 1141–1152. Printed in the UK PII: S0305-4470(98)88102-X

On the linear statistics of Hermitian random matrices

Yang Chen† and Nigel Lawrence
Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK

Received 2 October 1997

Abstract. In this paper we continue with the study of linear statistics of random Hermitian matrix ensembles. We generalize the result of a previous paper on the probability density function of linear statistics to the finite N ensembles where the interval of support of the eigenvalue spectrum is a single interval.

Combining the Hankel determinant formula for orthogonal polynomials which are associated with Hermitian matrix ensembles and the linear statistics theorem for finite N, we obtain the strong or oscillatory asymptotics for the polynomials orthogonal with respect to weight functions supported on the real axis.

1. Introduction

In the application of the theory of random matrices, for example in the theory of quantum transport in disordered systems [3], one is often encountered with the random variable

$$Q := \operatorname{tr} f(M) \tag{1.1}$$

where $f(M)$ is a real-valued function of the $N \times N$ matrix M.

The space of matrices has the probability measure [5]

$$\operatorname{prob}(M)\mathrm{d}M := \exp[-\operatorname{tr} v(M)]\,\mathrm{d}M = \operatorname{vol}(\beta, N) \prod_{1 \leqslant j < k \leqslant N} |x_j - x_k|^\beta \prod_{1 \leqslant l \leqslant N} \mathrm{e}^{-v(x_l)}\,\mathrm{d}x_l. \tag{1.2}$$

Here $\{x_j : 1 \leqslant j \leqslant N\}$ are the eigenvalues, $\beta = 1, 2, 4$ are for matrices with orthogonal, unitary and symplectic symmetries respectively and $\operatorname{vol}(\beta, N)$ are the corresponding volumes of the symmetry groups that diagonalize the respective matrices. For the purpose of this paper we shall assume that $v(x)$ is convex, for $x \in \mathbb{R}$, and therefore $v''(x)$ is positive on a set of positive measure.

2. Linear statistics

We suppose the probability density function of Q, denoted by $\mathcal{P}(Q)$, is known and compute its Fourier transform,

$$\hat{\mathcal{P}}(k) := \int_{-\infty}^{+\infty} \mathrm{d}Q\mathrm{e}^{-\mathrm{i}kQ}\mathcal{P}(Q) \tag{2.1}$$

where

$$\mathcal{P}(Q) := \langle \delta(Q - \operatorname{tr} f(M)) \rangle_M. \tag{2.2}$$

† E-mail address: y.chen@ic.ac.uk

The symbol $\langle\cdot\rangle_M$ denotes an average over the matrices

$$\langle g(M)\rangle_M := \frac{\int \mathrm{d}M \exp[-N \operatorname{tr} v(M)] g(M)}{\int \mathrm{d}M \exp[-N \operatorname{tr} v(M)]}. \tag{2.3}$$

Note that we have introduced the factor N in front of $v(M)$ for the convenience of evaluating the multiple integrals in a large N or Coulomb fluid approximation [8]. In what follows we shall only deal with the Hermitian case, i.e. $\beta = 2$.

Expressing dM in the eigenvalue form using (1.2), we have, writing ik as λ,

$$\hat{\mathcal{P}}(-\mathrm{i}\lambda) = \frac{Z_N(\lambda)}{Z_N(\lambda = 0)} := \exp[-(F_N(\lambda) - F_N(\lambda = 0))] \tag{2.4}$$

where

$$Z_N(\lambda) := \left(\prod_{j=1}^{N} \int_{-\infty}^{\infty} \mathrm{d}x_j\right) \exp\left[-\Phi(x_1, \ldots, x_N) - \lambda \sum_{n=1}^{N} f(x_n)\right]. \tag{2.5}$$

If we interpret x_j, $j = 1, \ldots, N$ as the positions of N charged particles (all carrying identical charges), then

$$\Phi(x_1, \ldots, x_N) := -2 \sum_{1\leqslant j<k\leqslant N} \ln|x_j - x_k| + N \sum_{j=1}^{N} v(x_j) \tag{2.6}$$

becomes the total energy of N repelling classical charged particles confined by a common external potential $Nv(x)$. The linear statistics, $f(x)$, becomes a perturbation to the original system. In the limit of large N the collection of particles can be approximated as a continuous fluid with a density $\sigma(\cdot)$ supported in J (a subset of $\mathbb{R}$). If σ is normalized to unity, then σ can be obtained as the solution of the following constrained minimization problem:

$$\min_\sigma F[\sigma] \qquad \text{subject to } \int_J \mathrm{d}x\, \sigma(x) = 1 \tag{2.7}$$

where

$$F(\lambda) := \int_J \mathrm{d}x\, \sigma(x)[N^2 v(x) + \lambda N f(x)] - N^2 \int_J \mathrm{d}x \int_J \mathrm{d}y\, \sigma(x) \ln|x - y| \sigma(y). \tag{2.8}$$

Here $F(\lambda)$ is interpreted as the free energy of the system under an external perturbation with 'strength' λ and $F(\lambda = 0)$ the free energy of the original system. Therefore, the Fourier transform of the linear statistics can be expressed as the change in the free energy due to perturbation;

$$\hat{\mathcal{P}}(\lambda) = \exp[-(F(\lambda) - F(\lambda = 0))]. \tag{2.9}$$

Upon minimization $\sigma(x)$ is found to satisfy the integral equation

$$N^2 v(x) + \lambda N f(x) - 2N^2 \int_J \mathrm{d}y \ln|x - y| \sigma(y) = NA \qquad x \in J \tag{2.10}$$

where A is a constant for $x \in J$ and NA is the Lagrange multiplier that fixes the constraint also known as the chemical potential. Differentiating (2.10) with respect to x gives the singular integral equation

$$N^2 v'(x) + \lambda N f'(x) - 2N^2 P \int_J \frac{\mathrm{d}y}{x - y} \sigma(y) = 0 \qquad x \in J \tag{2.11}$$

satisfied by σ. To solve for σ we write

$$\sigma(x) = \sigma_0(x) + \frac{\varrho(x)}{N} \tag{2.12}$$

with

$$\int_J \mathrm{d}x\, \sigma_0(x) = 1 \qquad \int_J \mathrm{d}x\, \varrho(x) = 0. \tag{2.13}$$

We suppose σ_0 solves,

$$v'(x) - 2P\int_J \frac{\mathrm{d}y}{x-y}\sigma_0(y) = 0 \tag{2.14}$$

and ϱ solves

$$\lambda f'(x) - 2P\int_J \frac{\mathrm{d}y}{x-y}\varrho(y) = 0. \tag{2.15}$$

We now discuss the solution for σ_0. Under the assumption that $v(x)$ is convex and consequently $v''(x) > 0$ in a set of positive measure, it can be shown that [2] $\sigma_0(x)$ is supported in a single interval (a, b). The solution subject to the boundary condition $\sigma_0(a) = 0 = \sigma_0(b)$ reads, according to the theory of singular integral equations [9],

$$\sigma_0(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi^2} P\int_a^b \frac{\mathrm{d}y}{y-x}\frac{v'(y)}{\sqrt{(b-y)(y-a)}} \qquad y \in (a, b) \tag{2.16}$$

with a supplementary condition,

$$\int_a^b \mathrm{d}x\, \frac{v'(x)}{\sqrt{(b-x)(x-a)}} = 0. \tag{2.17}$$

The normalization condition, $\int_a^b \mathrm{d}x\, \sigma(x) = \int_a^b \mathrm{d}x\, \sigma_0(x) = 1$, becomes,

$$\frac{1}{2\pi}\int_a^b \mathrm{d}x\, \frac{xv'(x)}{\sqrt{(b-x)(x-a)}} = 1. \tag{2.18}$$

The unique solution for ϱ subject to $\int_J \mathrm{d}x\, \varrho(x) = 0$, is

$$\varrho(x) = \frac{\lambda}{2\pi^2\sqrt{(b-x)(x-a)}} P\int_a^b \mathrm{d}y\, \frac{\sqrt{(b-y)(y-a)}}{y-x} f'(y) \tag{2.19}$$

recalling that the support of σ_0, J, is (a, b). Thus the change in free energy is

$$\begin{aligned} F(\lambda) - F(\lambda = 0) &= \frac{\lambda}{2}\int_a^b f(x)\varrho(x) + \lambda N\int_a^b \mathrm{d}x\, f(x)\sigma_0(x) \\ &= \frac{\lambda^2}{4\pi^2}\int_a^b \mathrm{d}x\, \frac{f(x)}{\sqrt{(b-x)(x-a)}} P\int_a^b \mathrm{d}y\, \frac{\sqrt{(b-y)(y-a)}}{y-x} f'(y) \\ &\quad + \lambda N\int_a^b \mathrm{d}x\, f(x)\sigma_0(x). \end{aligned} \tag{2.20}$$

Therefore, $\mathcal{P}(Q)$, the probability density of the linear statistics $f(x)$, is a Gaussian with mean,

$$N\int_a^b \mathrm{d}x\, f(x)\sigma_0(x) \tag{2.21}$$

and variance

$$\frac{1}{2\pi^2}\int_a^b \mathrm{d}x\, \frac{f(x)}{\sqrt{(b-x)(x-a)}} P\int_a^b \mathrm{d}y\, \frac{\sqrt{(b-y)(y-a)}}{x-y} f'(y). \tag{2.22}$$

If $\hat{\mathcal{P}}(k)$ is interpreted as the probability density *functional* of the local density random field, then the covariance or the two-point function of the local density (normalized so that the total number of eigenvalues is N) is

$$\mathrm{Cov}(x, y) = \frac{1}{2\pi^2} P \frac{ab + xy - (x+y)(a+b)/2}{\sqrt{(b-x)(x-a)}\sqrt{(b-y)(y-a)}(x-y)^2}. \qquad (2.23)$$

Note that in this normalization of the density, the right-hand side of (2.18) should be replaced by N. For convex $v(x)$, we have $b \to +\infty$, as $N \to \infty$, and $a \to -\infty$, as $N \to \infty$. Therefore,

$$\lim_{N\to\infty} \mathrm{Cov}(x, y) = -\frac{1}{2\pi^2(x-y)^2} \qquad (2.24)$$

reducing to a well known result [5].

In the next two sections, we give applications of the linear statistics formulae (2.21) and (2.22). In section 3 we give the probability density function for the number statistics, $n[A, B]$, defined to be the number of eigenvalues in an interval $[A, B]$ contained in $[a, b]$. This was first investigated in [3] where the eigenvalues are supported in $[0, \infty)$ and in the limit $N \to \infty$. We shall see that the probability density of the number of eigenvalues in an interval is a Gaussian with a universal variance behaving in a qualitatively similar way as that studied in [3]. In section 4, an interesting application of (2.20) is made to compute the strong asymptotics of the orthogonal polynomials based on the Hankel determinant formula. This gives the strong asymptotics of the polynomials in terms of the confining potential $v(x)$. Using the asymptotic formula we compute the reproducing kernel and investigate its bulk scaling form. By specializing $v(x)$ we establish a conjecture of Nevai on the Freud polynomials [6], where $v(x) = |x|^\alpha, x \in (-\infty, \infty), \alpha > 1$. This is found to be agreement with the result of [7]. In section 5, we indicate how the asymptotics of the gap orthogonal polynomials can be determined.

3. The number statistics $n[A, B]$

In this case the appropriate linear statistics is

$$f(x) := \chi_{[A,B]}(x) \qquad a < A \leqslant x \leqslant B < b \qquad (3.1)$$

where $\chi_{[A,B]}(x)$ is the characteristic function of the interval $[A, B]$. Note that the characteristic function has the convenient form, $\chi_{[A,B]}(x) = \frac{1}{2}[\theta(x-A) - \theta(A-x) - (\theta(x-B) - \theta(B-x))]$. Using (2.20) we find

$$F(\lambda) - F(\lambda = 0) = P\int_A^B \frac{\lambda^2\,\mathrm{d}x/(4\pi^2)}{\sqrt{(b-x)(x-a)}} \left[\frac{\sqrt{(b-A)(A-a)}}{A-x} - \frac{\sqrt{(b-B)(B-a)}}{B-x}\right] + \lambda N \int_A^B \mathrm{d}x\, \sigma_0(x) \qquad (3.2)$$

where the first integral is found by using the fact that $\chi'_{[A,B]}(x) = \delta(x-A) - \delta(x-B)$. The second integral is the average number of particles/eigenvalues in the interval $[A, B]$.

Denoting the first integral as $S[A, B]$, it can be seen in the bulk scaling limit, where $b \gg B$ and $|a| \gg |A|$, that

$$S[A, B] = -\mathrm{O}\left[\ln\left(\frac{B-A}{\varepsilon}\right)\right] \qquad B - A \gg \varepsilon \qquad (3.3)$$

where ε is the short distance cut-off dictated by the principal-value integral. Explicitly $S[A, B]$, can be expressed in terms of elementary functions, however, the logarithmic behaviour is not apparent.

To see this we specialize to even $v(x)$, and consequently $a = -b$. We also choose a symmetric interval for the characteristic function; $A = -B$. Thus, denoting $s = 2B$,

$$S[s] = -\frac{\lambda^2}{2\pi^2} \ln\left[\frac{s}{\varepsilon}\left(1 - \frac{s^2}{4b^2}\right)\right] \sim -\frac{\lambda^2}{2\pi^2} \ln\left(\frac{s}{\varepsilon}\right) \qquad b \gg s \gg \varepsilon. \tag{3.4}$$

Therefore the probability density of the number of eigenvalues in a symmetric interval of length s, confined by an even convex potential is a Gaussian centred at

$$N \int_{-s/2}^{s/2} \mathrm{d}x\, \sigma_0(x) \tag{3.5}$$

with variance

$$\frac{1}{\pi^2} \ln\left(\frac{s}{\varepsilon}\right). \tag{3.6}$$

This generalizes the result obtained in [3].

4. Strong asymptotics

In the theory of Hermitian random matrices, a fundamental quantity, denoted as $E[J]$, is the probability that the interval J of the spectrum has no eigenvalues. This quantity can be expressed as the Fredholm determinant of a certain integral operator, $\hat{K}$, over J [5]; $E[J] = \det[1 - \hat{K}_J]$, where $\hat{K}$ has kernel

$$K_N(x, y) = \sqrt{w(x)w(y)}\sqrt{\beta_N}\, \frac{\hat{p}_N(x)\hat{p}_{N-1}(y) - \hat{p}_N(y)\hat{p}_{N-1}(x)}{x - y}.$$

Here $w(x) = \mathrm{e}^{-v(x)}$ is weight function of the orthonormal polynomials $\hat{p}_n(x)$;

$$\int_{-\infty}^{+\infty} \mathrm{d}x\, w(x)\hat{p}_m(x)\hat{p}_n(x) = \delta_{m,n}$$

and satisfies the recurrence relations

$$x\,\hat{p}_n(x) = \sqrt{\beta_{n+1}}\,\hat{p}_{n+1}(x) + \alpha_n \hat{p}_n(x) + \sqrt{\beta_n}\,\hat{p}_{n-1}(x)$$

with $\alpha_n \in \mathbb{R}$ and $\beta_n > 0$. $K_N(x, y)$ is also known as the reproducing kernel.

It is therefore of interest to determine asymptotics for the polynomials in the bulk scaling limit, which corresponds to fixing x in the oscillatory region of the polynomials and with N large.

According to the Hankel determinant representation for monic polynomials $p_N(t)$ orthogonal with respect to the weight function $w(t)$, has the multiple-integral representation [10];

$$p_N(t) = \langle \det(t\mathrm{I} - M)\rangle_M = \frac{\left(\prod_{j=1}^N \int_{-\infty}^{+\infty} \mathrm{d}x_j\, w(x_j)\right) \prod_{1\leqslant l<n\leqslant N} |x_l - x_n|^2 \prod_{k=1}^N (t - x_k)}{\left(\prod_{j=1}^N \int_{-\infty}^{+\infty} \mathrm{d}x_j\, w(x_j)\right) \prod_{1\leqslant l<n\leqslant N} |x_l - x_n|^2}. \tag{4.1}$$

From this we recognize that the computation $p_N(t)$ is a special case of the linear statistics theorem, with $f(x; t) := -\ln(t - x)$, and $\lambda = 1$. It is well known that an orthogonal polynomial of degree N has N real simple zeros which are bounded by the end points of the support of σ_0, the zero counting function, denoted as a and b. It can be shown that $|a|$ and b are increasing functions of N if $v(x)$ is convex and are determined by

$$\int_a^b \mathrm{d}x\, \frac{v'(x)}{\sqrt{(b-x)(x-a)}} = 0 \qquad \int_a^b \mathrm{d}x\, \frac{xv'(x)}{\sqrt{(b-x)(x-a)}} = 2\pi N.$$

Note that here $\int_a^b \mathrm{d}x\, \sigma_0(x) = N$. The largest zero for large N is slightly smaller then b for convex $v(x)$, [2]. We therefore compute the change in the free energy for $t > b$, so that $\ln(t-x) > 0$, for $x \in (a, b)$. It can be shown that the formula given below is valid for t outside the interval in which the zeros of $p_N(t)$ are contained. For $t > b$ and $t < a$, $p_N(t)$ does not oscillate. Thus denoting the change in the free energy by $S(t)$, gives,

$$p_N(t) = \exp[-S(t)] \qquad t \notin [a, b] \tag{4.2}$$

where

$$S(t) = S_1(t) + S_2(t) \tag{4.3}$$

with

$$S_1(t) := \frac{1}{4\pi^2} \int_a^b \mathrm{d}x\, \frac{\ln(t-x)}{\sqrt{(b-x)(x-a)}} P \int_a^b \mathrm{d}y\, \frac{\sqrt{(b-y)(y-a)}}{(y-x)(y-t)} \tag{4.4}$$

$$S_2(t) := -\int_a^b \mathrm{d}x\, \ln(t-x)\sigma_0(x) \tag{4.5}$$

for $t \notin [a, b]$. To determine the strong asymptotics, for $t \in [a, b]$, where $p_N(t)$ oscillates, we define $p_N(t), t \in [a, b]$ to be the real part of the analytic continuation of $p_N(t), t \notin [a, b]$. Thus

$$p_N(t) := \Re[\exp[-S(t + \mathrm{i}\varepsilon)]] \qquad t \in [a, b] \qquad \varepsilon \to 0. \tag{4.6}$$

We find after some elementary computations,

$$S_1(t) = \frac{1}{4} \ln\left[\frac{16(t-a)(t-b)}{(b-a)^2} \left[\frac{\sqrt{t-a} - \sqrt{t-b}}{\sqrt{t-a} + \sqrt{t-b}}\right]^2\right] \qquad t \notin [a, b] \tag{4.7}$$

$$S_2(t) = -N \ln\left(\frac{\sqrt{t-a} + \sqrt{t-b}}{2}\right)^2 + \int_a^b \frac{\mathrm{d}y}{2\pi} \frac{v(y)}{\sqrt{(b-y)(y-a)}} \left[\frac{\sqrt{(t-a)(t-b)}}{y-t} + 1\right] \qquad t \notin [a, b] \tag{4.8}$$

where the normalization and supplementary conditions have been used to arrive at $S_2(t)$.

The behaviour of $S_1(t)$ for $\Re t \in [a, b]$, and $\Im t = \varepsilon \to 0$, can be found by the parametrization;

$$t - a = |t-a|\mathrm{e}^{\mathrm{i}\theta_a} \qquad t - b = |t-b|\mathrm{e}^{\mathrm{i}\theta_b} \qquad \theta_a = \varepsilon \qquad \theta_b = \pi - \varepsilon.$$

Thus

$$\mathrm{e}^{-S_1(t+\mathrm{i}\varepsilon)} = \sqrt{\frac{b-a}{4}} \frac{1}{[(t-a)(b-t)]^{1/4}} \exp\left[\mathrm{i}\left(\phi(t) - \frac{\pi}{4}\right)\right] \qquad t \in [a, b] \tag{4.9}$$

and the angle $\phi(t)$ can be parametrized in several ways:

$$\cos[2\phi(t)] = \frac{2t - (b+a)}{b-a} \qquad \tan[\phi(t)] = \frac{(b-t)^{1/2}}{(t-a)^{1/2}}. \tag{4.10}$$

To determine $S_2(t+\mathrm{i}\varepsilon), t \in [a,b]$, we use the same procedure as in determining $S_1(t+\mathrm{i}\varepsilon)$ and the Sokhotski–Plemelj formula [9] and find

$$\mathrm{e}^{-S_2(t+\mathrm{i}\varepsilon)} = \left(\frac{\sqrt{b-a}}{2}\right)^{2N} \exp\left[\frac{v(t)}{2} - \int_a^b \frac{\mathrm{d}x}{2\pi}\frac{v(x)}{\sqrt{(b-x)(x-a)}}\right]$$
$$\times \exp\left[2\mathrm{i}N\phi(t) - \mathrm{i}P\int_a^b \frac{\mathrm{d}x}{2\pi}\frac{v(x)}{x-t}\frac{\sqrt{(b-t)(t-a)}}{\sqrt{(b-x)(x-a)}}\right] \qquad t \in [a,b].$$

Using the above information, we find that the orthogonal polynomials have the following strong asymptotics expansion,

$$\sqrt{w(t)}\,p_N(t) \sim \left(\frac{\sqrt{b-a}}{2}\right)^{2N+1} \frac{1}{[(b-t)(t-a)]^{1/4}} \exp\left[-\int_a^b \frac{\mathrm{d}x}{2\pi}\frac{v(x)}{\sqrt{(b-x)(x-a)}}\right]$$
$$\times \cos\left[(2N+1)\phi(t) - \frac{\pi}{4} - P\int_a^b \frac{\mathrm{d}x}{2\pi}\frac{v(x)}{x-t}\frac{\sqrt{(b-t)(t-a)}}{\sqrt{(b-x)(x-a)}}\right]$$
$$t \in [a,b].$$

This gives the strong asymptotics expansion directly in terms of the confining potential, $v(x)$.

There is an alternative evaluation of $S_2(t)$, which has the advantage of exposing the parity rule satisfied by $p_N(t)$ for even $v(x)$, $p_N(-t) = (-1)^N p_N(t), t \in \mathbb{R}$. First we compute $\frac{\mathrm{d}S_2}{\mathrm{d}t}, t \notin [a,b]$. Thus

$$S_2'(t) = -\int_a^b \mathrm{d}x\,\frac{\sigma_0(x)}{t-x} \qquad \text{and} \qquad S_2'(t)\Big|_{t\to t+\mathrm{i}\varepsilon} = -P\int_a^b \mathrm{d}x\,\frac{\sigma_0(x)}{t-x} + \mathrm{i}\pi\sigma_0(t)$$
$$= -\frac{v'(t)}{2} + \mathrm{i}\pi\sigma_0(t).$$

Integrating from b to t, gives

$$S_2(t+\mathrm{i}\varepsilon) = S_2(b) - \frac{[v(t)-v(b)]}{2} + \mathrm{i}\pi\int_b^t \mathrm{d}s\,\sigma_0(s) \qquad t \in [a,b].$$

Putting $x = b$ in the integral equation satisfied by σ_0,

$$v(x) - 2\int_a^b \mathrm{d}y\ \ln|x-y|\sigma_0(y) = A$$

we find

$$\frac{v(b)}{2} - \int_a^b \mathrm{d}y\ \ln|b-y|\sigma_0(y) = \frac{A}{2}.$$

From the definition of $S_2(t)$,

$$\frac{v(b)}{2} + S_2(b) = \frac{A}{2}.$$

Thus

$$\mathrm{e}^{-S_2(t+\mathrm{i}\varepsilon)} = \exp\left[-\frac{A}{2} + \frac{v(t)}{2} - \mathrm{i}\pi\int_b^t \mathrm{d}x\,\sigma_0(x)\right] \qquad t \in [a,b]$$

and we have the alternative asymptotics expansion,

$$\sqrt{w(t)}\,p_N(t) \sim \sqrt{\frac{b-a}{4}}\,\frac{\mathrm{e}^{-A/2}}{[(t-a)(b-t)]^{1/4}} \cos\left[\phi(t) - \frac{\pi}{4} - \pi\int_b^t \mathrm{d}x\,\sigma_0(x)\right]. \qquad (4.11)$$

If $v(x)$ is even, then $a = -b$ and $\phi(-t) = -\phi(t) + \pi/2$. Furthermore, since $\sigma_0(x)$ is even, we have

$$\int_b^{-t} \mathrm{d}x\, \sigma_0(x) = -N + \int_t^b \mathrm{d}x\, \sigma_0(x).$$

Using these the parity rule satisfied by $p_N(t)$ is easily established. We can also obtain the orthonormal polynomials through $\hat{p}_N(t) = p_N(t)/\sqrt{h_N}$, where h_N is the square of the L^2 norm of $p_N(t)$. However, from standard theory, see for example [2],

$$h_N = \exp[-(F_{N+1} - F_N)] \sim \exp\left[-\frac{\partial F}{\partial N}\right] = \exp[-A]. \tag{4.12}$$

Thus

$$\hat{p}_N(t) = \exp\left[\frac{A}{2}\right] p_N(t). \tag{4.13}$$

We note here an interesting identity relating A and a certain integral involving the external potential $v(x)$:

$$\frac{A}{2} - \int_a^b \frac{\mathrm{d}x\, v(x)}{2\pi\sqrt{(b-x)(x-a)}} = -N \ln\left(\frac{b-a}{4}\right).$$

Thus

$$[(b-t)(t-a)]^{1/4}\sqrt{w(t)}\,\hat{p}_N(t) \sim \sqrt{\frac{b-a}{4}} \cos[\Psi_N(t)] \qquad t \in [a, b] \tag{4.14}$$

where

$$\begin{aligned}\Psi_N(t) &= (2N+1)\phi(t) - \frac{\pi}{4} - P\int_a^b \frac{\mathrm{d}x}{2\pi}\frac{v(x)}{x-t}\frac{\sqrt{(b-t)(t-a)}}{\sqrt{(b-x)(x-a)}} \\ &= \phi(t) - \frac{\pi}{4} - \pi\int_b^t \mathrm{d}x\, \sigma_0(x).\end{aligned} \tag{4.15}$$

Recognizing that $\beta_N = (b-a)^2/16 + \mathrm{O}(N^{-\mu})$, $\mu > 0$ [2], we have

$$\sqrt{w(t)}\left[\frac{(b-t)(t-a)}{\beta_N}\right]^{1/4} \hat{p}_N(t) = c\cos[\Psi_N(t)] + \mathrm{o}(1) \tag{4.16}$$

where c is constant to be determined by the approximate normalization condition on the polynomials,

$$\int_a^b \mathrm{d}t\, w(t)[\hat{p}_N(t)]^2 = 1.$$

Using the asymptotic formula (4.16), and replacing $\cos^2[\Psi_N(t)]$ by its root-mean-square value, $\frac{1}{2}$, we find

$$c = 2\sqrt{\frac{2}{\pi(b-a)}}.$$

Thus the normalized polynomials read, for $t \in [a, b]$,

$$\sqrt{w(t)}[(b-t)(t-a)]^{1/4}\hat{p}_N(t) = \sqrt{\frac{2}{\pi}}\cos[\Psi_N(t)] + \mathrm{o}(1). \tag{4.17}$$

From the strong asymptotics of the orthonormal polynomials, the reproducing kernel reads,

$$K_N(t_1, t_2) = \frac{\cos[\Psi_N(t_1)]\cos[\Psi_{N-1}(t_2)] - \cos[\Psi_N(t_2)]\cos[\Psi_{N-1}(t_1)]}{\pi(t_1 - t_2)\sqrt{\sin[2\phi(t_1)]}\sqrt{\sin[2\phi(t_2)]}} \tag{4.18}$$

where we have assumed that

$$b(N+1)-b(N)=\mathrm{o}(b(N)) \qquad \text{and} \qquad a(N+1)-a(N)=\mathrm{o}(a(N))$$

which is justified if $v(x)$ has polynomial increase for sufficiently large x, [2]. The reproducing kernel can be further transformed into a useful form for the 'bulk-scaling' limit, to be described later.

For sufficiently large N,

$$\Psi_{N-1}(t)=\phi(t)-\frac{\pi}{4}-\pi\int_b^t \mathrm{d}x\,\sigma_0(x;N-1)+\mathrm{o}(1)$$

and (see [2])

$$\sigma_0(x,N-1)=\sigma_0(x,N)-\frac{\partial\sigma_0(x,N)}{\partial N}+\mathrm{o}(1).$$

Also, for $x\in[a,b]$, (see [2])

$$\frac{\partial\sigma_0(x,N)}{\partial N}=\frac{1}{\pi\sqrt{(b-x)(x-a)}}.$$

Using these,

$$\begin{aligned}\Psi_{N-1}(t)&=\Psi_N(t)+\int_b^t\frac{\mathrm{d}x}{\sqrt{(b-x)(x-a)}}+\mathrm{o}(1)\\&=\Psi_N(t)-2\phi(t)+\mathrm{o}(1).\end{aligned}$$

The reproducing kernel becomes,

$$\begin{aligned}K_N(t_1,t_2)&=\frac{\cos[\Psi_N(t_1)]\cos[\Psi_N(t_2)-2\phi(t_2)]-\cos[\Psi_N(t_2)]\cos[\Psi_N(t_1)-2\phi(t_1)]}{\pi(t_1-t_2)\sqrt{\sin[2\phi(t_1)]}\sqrt{\sin[2\phi(t_2)]}}\\&=-\frac{\sin[\eta(t_1)+\eta(t_2)]\sin[\phi(t_1)-\phi(t_2)]+\sin[\eta(t_1)-\eta(t_2)]\sin[\phi(t_1)+\phi(t_2)]}{\pi(t_1-t_2)\sqrt{\sin[2\phi(t_1)]}\sqrt{\sin[2\phi(t_2)]}}\end{aligned} \tag{4.19}$$

where

$$\eta(t):=-\frac{\pi}{4}-\pi\int_b^t\mathrm{d}x\,\sigma_0(x)\qquad t\in[a,b].$$

By the taking the limit $t_1\to t_2$, we find

$$K_N(t,t)=\sigma_0(t)-\frac{b-a}{4\pi(b-t)(t-a)}\cos\left[2\pi\int_b^t\mathrm{d}x\,\sigma_0(x)\right].$$

The second term of the previous equation gives an oscillatory correction to the Coulomb fluid density. Numerically, using Mathematica, $K_N(t,t)$, computed for the example of the Hermite polynomials, agrees very well with that obtained from the strong asymptotic formula, except near the end points.

In the bulk-scaling limit, where $(b-a)\to\infty$ and $|t_1-t_2|\ll(b-a)$, it is clear that

$$\sin(\phi(t_1)-\phi(t_2))=\mathrm{O}\left(\frac{|t_1-t_2|}{b-a}\right) \tag{4.20}$$

while the denominator of (4.19) is $\mathrm{O}(|t_1-t_2|)$. Therefore,

$$K_N^{\mathrm{bulk}}(t_1,t_2)\approx\frac{\sin[\pi\int_{t_2}^{t_1}\mathrm{d}x\,\sigma_0(x)]}{\pi(t_1-t_2)}. \tag{4.21}$$

As an example of the application of (4.17), consider the Freud weight; $v(x) = |x|^\alpha, \alpha > 1$, with $t = b\cos\theta, \theta \in (0, \pi)$. The principal-value integral in (4.15) can be expressed in terms of a particular hypergeometric function;

$$\Psi_N(b\cos\theta, \alpha) = N[\theta - \cos\theta\sin\theta\,{}_2F_1(1-\alpha/2, 1; \tfrac{3}{2}; \sin^2\theta)] + \theta/2 - \pi/4.$$

Thus

$$\sqrt{b\sin\theta w(b\cos\theta)}\,\hat{p}_N(b\cos\theta) = \sqrt{\frac{2}{\pi}}\cos[\Psi_N(b\cos\theta, \alpha)] + \mathrm{o}(1) \tag{4.22}$$

where b is related to the degree N;

$$b^\alpha = \frac{\Gamma^2(\alpha/2)2^{\alpha-1}N}{\Gamma(\alpha)}[1 + \mathrm{o}(1)] \tag{4.23}$$

which establishes a conjecture of [6].

5. Generalization to multi-intervals

Suppose $v(x, \{g\})$ is not convex, has polynomial increase near infinity and has local minima separated by local maxima. Here $\{g\} = (g_1, g_2, \ldots)$, is a finite set of real 'adjustable' constants. Based on the Coulomb fluid picture of [2], we see that a possible fluid/eigenvalue density, $\sigma_0(x)$, is that which minimizes $F[\sigma_0, \lambda = 0]$, is a solution of the integral equation

$$v(x) - 2\int_L \mathrm{d}y\ \ln|x-y|\sigma_0(y) = A \qquad x \in L \tag{5.1}$$

and which is supported in the union of m mutually disjoint intervals;

$$L = \cup_{j=1}^m L_j \qquad L_j = [a_j, b_j] \tag{5.2}$$

provided the constants, $\{g\}$, are chosen appropriately. We give a brief description of how this may be accomplished. We start with a single interval solution which vanishes at the end points of the interval. Suppose the set $\{g\}$ is tuned to a set of critical values $\{g^c\}$ for which the density vanishes at points contained in the interval (excluding the end points). If $\{g\}$ is increased beyond the critical values, then the density breaks up into 'lumps' supported in a union of disjoint intervals.

In this situation, the $\sigma_0(x)$ which vanishes at the end points of L reads [9]

$$\sigma_0(x) = \frac{\sqrt{R(x)}}{2\pi^2} P\int_L \mathrm{d}y \frac{v'(y)}{(y-x)\sqrt{R(y)}} \qquad x \in L \tag{5.3}$$

supplemented by m side conditions,

$$\int_{L_j} \frac{\mathrm{d}x\, x^{j-1}v'(x)}{\sqrt{R(x)}} = 0 \qquad 1 \leqslant j \leqslant m \tag{5.4}$$

where

$$R(x) := \prod_{j=1}^m (x - a_j)(x - b_j). \tag{5.5}$$

Note that (5.3) has the alternative form,

$$\sigma_0(x) = \frac{\sqrt{R(x)}}{2\pi^2}\int_L \frac{\mathrm{d}y}{\sqrt{R(y)}}\frac{v'(x) - v'(y)}{x - y}. \tag{5.6}$$

Let $\{x_j^{\min}, 1 \leqslant j \leqslant m\}$ denote the m local minima of $v(x)$ with a suitably chosen set $\{g\}$, then $a_j + \varepsilon < x_j^{\min} < b_j - \varepsilon$. If $v(x)$ is sectionally convex, i.e. $v''(x) > 0$, for $x \in (a_j, b_j)$, then

$$\lim_{x \to b_j^-} \frac{\sigma_0(x)}{\sqrt{b_j - x}} = G(a_1, , , a_m, b_1, , , b_m) > 0 \qquad 1 \leqslant j \leqslant m \tag{5.7}$$

and

$$\lim_{x \to a_j^+} \frac{\sigma_0(x)}{\sqrt{x - a_j}} = H(a_1, , , a_m, b_1, , , b_m) > 0 \qquad 1 \leqslant j \leqslant m. \tag{5.8}$$

This generalizes the one interval result obtained in [2].

Together with the normalization condition, $\int_L \mathrm{d}x\, \sigma_0(x) = N$, there are only $m + 1$ conditions; not enough to determine the $2m$ unknowns $\{a_j, b_j, 1 \leqslant j \leqslant m\}$. This problem can be solved if we could find extra side conditions to complete the existing ones, (5.4). From the above argument we see that the density $\sigma_0(x)$ 'breaks up' into m pieces each of which contains a fraction of the total number of particles, N. More precisely,

$$\int_{L_j} \mathrm{d}x\, \sigma_0(x) = \alpha_j N \qquad 0 \leqslant \alpha_j \leqslant 1 \qquad 1 \leqslant j \leqslant m. \tag{5.9}$$

Note that due to the normalization condition,

$$\alpha_m = 1 - \sum_{j=1}^{m-1} \alpha_j. \tag{5.10}$$

Therefore we have $2m$ conditions; (5.4) and (5.9) for inverting the $2m$ unknowns, $\{a_j, b_j, 1 \leqslant j \leqslant m\}$ in terms of $\{\alpha_1, , , \alpha_{m-1}, N\}$. In order to determine the parameters, $\alpha_j, 1 \leqslant j \leqslant m - 1$, in terms of N, we propose a supplementary minimum principle. Substituting the the fluid density, $\sigma_0(x)$, back into the free energy functional, we see that the free energy, F, is a function of $\{\alpha_j, 1 \leqslant j \leqslant m - 1, N\}$;

$$F = F(\alpha_1, , , \alpha_{m-1}, N). \tag{5.11}$$

Therefore the sought after $\{\alpha_j, 1 \leqslant j \leqslant m - 1\}$ is such (5.11) is minimized. The necessary condition reads,

$$\frac{\partial F}{\partial \alpha_j} = 0 \qquad 1 \leqslant j \leqslant m - 1. \tag{5.12}$$

We conclude this paper by generalizing the linear statistics formula (2.20) to the multiple interval case:

$$\hat{\mathcal{P}}(k, [f]) = \exp[-S[f]] \tag{5.13}$$

where

$$S[f] := \frac{k^2}{4\pi^2} \int_L \frac{\mathrm{d}x f(x)}{\sqrt{R(x)}} P \int_L \frac{\mathrm{d}y f'(y)\sqrt{R(y)}}{x - y} + \mathrm{i}k \int_L \mathrm{d}x\, f(x)\sigma_0(x) \tag{5.14}$$

where $\int_L \mathrm{d}x \sigma_0(x) = N$, and the end points are determined by (5.4), (5.9) and (5.12).

The application of (5.14) to the number statistics and the gap orthogonal polynomials will be made in a future publication.

For applications of the linear statistics formula, such as the computation of the conductance of a disordered systems, we refer the readers to the review article [1] and the references therein and to [3] where the probability density function of arbitrary linear statistics was first obtained in $N \to \infty$ limit. We note here that (1.2) with $g(M)$ specialized to $\det(\mathrm{I}t - M)$ and $v(M)$ specialised to M^2 can be computed using supersymmetric methods [4]. However, the methods introduced in this paper only requires that $v''(x) > 0$ and does not require $v(x) = x^2$.

References

[1] Beenakker C W J 1997 *Rev. Mod. Phys.* **69** 731–803

[2] Chen Y and Ismail M E H 1997 Thermodynamic relations of the Hermitian matrix ensembles *J. Phys. A: Math. Gen.* **30** 6633–54 and references therein

[3] Chen Y and Manning S 1994 Distribution of linear statistics in random matrix models *J. Phys.: Condens. Matter* **6** 3039–44

[4] Efetov K B 1997 *Supersymmetry in Disorder and Chaos* (Cambridge: Cambridge University Press)

[5] Mehta M L 1991 *Random Matrices* (New York: Academic)

[6] Nevai P 1986 Géza Freud, orthogonal polynomials and Christoffel functions: a case study *J. Approx. Theory* **48** 3–167

[7] Rakhmanov E A 1992 *Progress in Approximation Theory (Springer Series in Computational Mathematics)* vol 9, ed E B Saff and A A Gonchar

[8] Dyson F J 1962 Statistical theory of energy levels of complex systems I *J. Math. Phys.* **3** 140–156
Dyson F J 1962 Statistical theory of energy levels of complex systems II *J. Math. Phys.* **3** 157–165
Dyson F J 1962 Statistical theory of energy levels of complex systems III *J. Math. Phys.* **3** 166–175. See also [2] and the references therein

[9] Gakhov F D 1990 *Boundary Value Problems* (New York: Dover)

[10] Szegö G 1975 *Orthogonal Polynomials* 4th edn (Providence, RI: American Mathematical Society)