Large deviations of the maximum eigenvalue in Wishart random matrices

2007 J. Phys. A: Math. Theor. 40 4317

(http://iopscience.iop.org/1751-8121/40/16/005)

J. Phys. A: Math. Theor. **40** (2007) 4317–4337 doi:10.1088/1751-8113/40/16/005

Large deviations of the maximum eigenvalue in Wishart random matrices

Pierpaolo Vivo[1], **Satya N Majumdar**[2] **and Oriol Bohigas**[2]

[1] School of Information Systems, Computing & Mathematics, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK
[2] Laboratoire de Physique Théorique et Modèles Statistiques (UMR 8626 du CNRS), Université Paris-Sud, Bâtiment 100, 91405 Orsay Cedex, France

E-mail: pierpaolo.vivo@brunel.ac.uk

Received 16 January 2007, in final form 27 February 2007
Published 30 March 2007
Online at stacks.iop.org/JPhysA/40/4317

Abstract

We analytically compute the probability of large fluctuations to the left of the mean of the largest eigenvalue in the Wishart (Laguerre) ensemble of positive definite random matrices. We show that the probability that all the eigenvalues of a $(N \times N)$ Wishart matrix $W = X^T X$ (where X is a rectangular $M \times N$ matrix with independent Gaussian entries) are smaller than the mean value $\langle\lambda\rangle = N/c$ decreases for large N as $\sim\exp\left[-\frac{\beta}{2}N^2\Phi_-\left(\frac{2}{\sqrt{c}}+1;c\right)\right]$, where $\beta = 1, 2$ corresponds respectively to real and complex Wishart matrices, $c = N/M \leqslant 1$ and $\Phi_-(x;c)$ is a rate (sometimes also called large deviation) function that we compute explicitly. The result for the anti-Wishart case $(M < N)$ simply follows by exchanging M and N. We also analytically determine the average spectral density of an ensemble of Wishart matrices whose eigenvalues are constrained to be smaller than a fixed barrier. Numerical simulations are in excellent agreement with the analytical predictions.

PACS numbers: 02.50.−r, 02.10.Yn, 24.60.−k

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Consider a rectangular $(M \times N)$ matrix X whose elements X_{ij} represent some data. The N entries of each of the M rows constitute the components of an N-dimensional vector $\vec{X}_i$ (with $i = 1, 2, \ldots, M$). The vector $\vec{X}_i$ (the ith row of the array) represents the ith sample of the data and the matrix element X_{ij} represents the jth component of the vector $\vec{X}_i$. For example, suppose we are considering a population of M students in a class, and for each student we have the data of their heights, their marks in an examination, their weights, etc forming a

vector of N elements (or traits) for each of the M students. Then the product $W = X^T X$ is a positive definite symmetric $(N \times N)$ matrix that represents the covariance matrix of the data (unnormalized). This matrix characterizes the correlations between different traits. The spectral properties of this matrix, i.e. its eigenvectors and eigenvalues, play a very important role in the so-called 'principal components analysis' (PCA) of multivariate data, a technique that is used regularly in detecting hidden patterns in data and also in image processing [1–3], amongst other applications. In PCA, one diagonalizes the covariance matrix W and identifies all the eigenvalues and eigenvectors. The data are usually maximally scattered in the direction of its principal eigenvector, corresponding to the largest eigenvalue, and are least scattered in the direction of the eigenvector, corresponding to the minimum eigenvalue. One can then prune the data by successively getting rid of the components (setting them to zero) along the eigenvectors corresponding to the smaller eigenvalues, but retaining the components along the larger eigenvalues, in particular those corresponding to the maximal eigenvalue. This method thus reduces the effective dimension of the data. This technique is called 'dimension reduction' and forms the basis of, e.g, image compression in computer vision [3].

When the underlying data are random, e.g. the elements of the matrix X are independent and identically distributed (i.i.d) random variables, real or complex, drawn from a Gaussian distribution, the product matrices $W = X^\dagger X$ constitute the so-called Wishart ensemble, named after Wishart who first introduced them [4]. In the literature one can also find the term 'Laguerre' ensemble, because the Laguerre polynomials arise in the analytical treatment of its spectral properties.

These Wishart random matrices have been extremely useful in multivariate statistical data analysis [1, 5] mentioned above (where W represents the covariance matrix) with applications in various fields ranging from meteorological data [6] to finance [7, 8]. Such matrices are also useful to analyse the capacity of channels with multiple antennae and receivers [9]. They also appear in nuclear physics [10], quantum chromodynamics [11] and also in statistical physics such as in a class of $(1 + 1)$-dimensional directed polymer problems [12]. Recently, Wishart matrices have also been used in the context of knowledge networks [13] and new mathematical results for the distribution of the matrix elements for the anti-Wishart matrices (when $M < N$) have been obtained [14, 15].

Given that the joint distribution of the elements of the $(M \times N)$ matrix X (real or complex) is a Gaussian, $P[X] \propto \exp\left[-\frac{\beta}{2}\mathrm{tr}(X^\dagger X)\right]$, where the Dyson index $\beta = 1, 2$ corresponds respectively to real and complex matrices [16], the spectral properties of the Wishart matrix $W = X^\dagger X$ have been studied extensively for many decades. For the case when $M \geqslant N$ (the number of samples is larger than the dimension) it is known that all the eigenvalues are positive, a typical eigenvalue scales as $\lambda \sim N$ for large N and the average density of eigenvalues in the large N limit has a scaling form $\rho_N(\lambda) \approx \frac{1}{N} f\left(\frac{\lambda}{N}\right)$, where $f(x)$ is the Marčenko–Pastur [17] function on the compact support $x \in [x_-, x_+]$:

$$f(x) = \frac{1}{2\pi x}\sqrt{(x_+ - x)(x - x_-)} \tag{1}$$

with $x_\pm = \left(\frac{1}{\sqrt{c}} \pm 1\right)^2$ and $c = N/M$ (with $c \leqslant 1$). (This result was also rederived by a different method by Dyson [18], and the spectral fluctuations were numerically investigated by Bohigas and Flores [19]). Thus, for $c \leqslant 1$, all the eigenvalues lie within a compact Marčenko–Pastur sea and the average eigenvalue is

$$\langle\lambda\rangle = \int_0^\infty \rho_N(\lambda)\lambda\, \mathrm{d}\lambda = \frac{N}{c}. \tag{2}$$

For all $c < 1$, the distribution goes to zero at the edges x_- and x_+. For the case $c = 1$ ($x_- = 0$ and $x_+ = 4$), the distribution diverges as $x^{-1/2}$ at the origin, $f(x) = \frac{1}{2\pi}\sqrt{(4 - x)/x}$ for

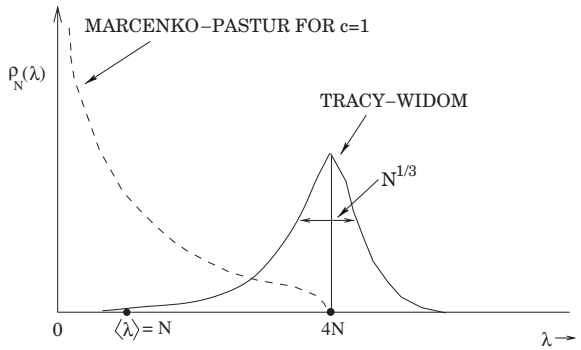

Figure 1. The dashed line schematically shows the Marčenko–Pastur form of the average density of states for $c = 1$. The average eigenvalue for $c = 1$ is $\langle\lambda\rangle = N$. For $c = 1$, the largest eigenvalue is centred around its mean $\langle\lambda_{\max}\rangle = 4N$ and fluctuates over a scale of width $N^{1/3}$. The probability of fluctuations on this scale is described by the Tracy–Widom distribution (shown schematically).

$0 \leqslant x \leqslant 4$ (shown schematically in figure 1). For the anti-Wishart case ($M < N$, i.e. $c > 1$), where one has M positive eigenvalues (the rest of the $(N - M)$ eigenvalues are identically zero), the corresponding result can be obtained from the $M \geqslant N$ case simply by exchanging M and N.

Another important issue in the context of PCA is the distribution of the largest eigenvalue of a Wishart random matrix and a lot of recent work has been devoted to this question [5, 12, 20–23]. From the exact analytical form of the density of states, it follows that the average of the maximum eigenvalue for large N is $\langle\lambda_{\max}\rangle \approx x_+(c)N$ where $x_+(c) = \left(\frac{1}{\sqrt{c}} + 1\right)^2$. However, for finite but large N, the maximum eigenvalue fluctuates, around its mean $x_+(c)N$, from one sample to another. A natural question is: what is the full probability distribution of the largest eigenvalue $\lambda_{\max}$? Recently, Johansson [12] and independently Johnstone [5] showed that for large N these fluctuations *typically* occur over a scale $\sim O(N^{1/3})$ around the mean, i.e. the upper edge of the Marčenko–Pastur distribution, and the probability of *typical* fluctuations $\chi = N^{-1/3}[\lambda_{\max} - x_+(c)N]$, properly centred and scaled, is described by the well-known Tracy–Widom distribution (see section 2 for details).

Note that the Tracy–Widom distribution describes the probability of *typical and small* fluctuations of $\lambda_{\max}$ over a narrow region of width $\sim O(N^{1/3})$ around the mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$. A question that is particularly important in the context of PCA is how to describe the probability of *atypical and large* fluctuations of $\lambda_{\max}$ around its mean, say over a wider region of width $\sim O(N)$? For example, what is the probability that all the eigenvalues of a Wishart random matrix are less than the average $\langle\lambda\rangle \approx N/c$ for large N? This is the same as the probability that $\lambda_{\max} \leqslant N/c$. Since $\langle\lambda_{\max}\rangle \approx x_+(c)N$, this requires the computation of the probability of an extremely rare event characterizing a large deviation of $\sim O(N)$ to the left of the mean (see e.g. a schematic picture for $c = 1$ in figure 1). Questions of this kind have been recently addressed in [24] on which we heavily rely, while for the general large deviations theory in connection with random matrices the reader is referred to [25].

In the context of PCA, this *large deviation* issue arises quite naturally because one is there interested in getting rid of redundant data by the 'dimension reduction' technique and keeping only the principal part of the data in the direction of the eigenvector representing the maximum eigenvalue, as mentioned before. The 'dimension reduction' technique works efficiently only if the largest eigenvalue is much larger than the other eigenvalues. However, if the largest eigenvalue is comparable to the average eigenvalue $\langle\lambda\rangle$, the PCA technique is not very useful.

Thus, the knowledge of large negative fluctuations of $\lambda_{\max}$ from its mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$ provides useful information about the efficiency of the PCA technique.

The main purpose of this paper is to provide exact analytical results for these large negative fluctuations of $\lambda_{\max}$ from its mean value. Rigorous mathematical results about the asymptotics of the Airy-kernel determinant (i.e. the probability that the largest eigenvalue lies deep inside the Marčenko–Pastur sea) for the cases $c = 1$ and $\beta = 2$ have been recently obtained [26]. Here we follow the Coulomb gas approach [16, 27], which interprets the eigenvalues of a random matrix as a fluid of charged interacting particles, and use standard functional integration methods of statistical physics. This approach has been exploited in the context of the Laguerre ensemble for the first time by Chen and Manning [28], who performed a detailed asymptotic analysis of the level spacing for general β and $c > 0$ and determined the distribution of the two smallest eigenvalues' in a certain double-scaling limit. These techniques have been also recently used to obtain analytically the large negative fluctuations of the maximum eigenvalue for the Gaussian ensembles [24]. Here we adopt this method for the Wishart ensemble.

We show that for $c \leqslant 1$, the probability of large fluctuations to the left of the mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$ behaves, for large N, as

$$\mathrm{Prob}[\lambda_{\max} \leqslant t, N] \sim \exp\left[-\frac{\beta}{2}N^2\Phi_-\left(\frac{x_+(c)N - t}{N}; c\right)\right], \tag{3}$$

where $t \sim O(N) \leqslant x_+(c)N$ is located deep inside the Marčenko–Pastur sea and $\Phi_-(x; c)$ is a certain *left* rate (sometimes also called large deviation) function with x being the main argument of the function and c being a parameter. In this paper, we compute the rate function $\Phi_-(x; c)$ explicitly. Knowing this function, it then follows that for large N

$$\mathrm{Prob}[\lambda_{\max} \leqslant \langle\lambda\rangle = N/c, N] \sim \exp(-\theta(c)N^2), \tag{4}$$

where the coefficient

$$\theta(c) = \frac{\beta}{2}\Phi_-\left(\frac{2}{\sqrt{c}} + 1; c\right). \tag{5}$$

For example, for the case $c = 1$ ($M = N$), we show that

$$\theta(1) = \beta\left(\log 2 - \frac{33}{64}\right) = 0.177\,522\ldots\beta. \tag{6}$$

The corresponding result for the anti-Wishart matrices ($M < N$) simply follows by exchanging M and N. In this paper, we focus only on the *left* large deviations of $\lambda_{\max}$. The corresponding probability of large fluctuations of $\lambda_{\max}$ to the *right* of the mean $\langle\lambda_{\max}\rangle$ was previously computed explicitly by Johansson [12] (see the following section for details).

As a byproduct of our analysis, we provide the general expression for the spectral density of a constrained Wishart ensemble of matrices whose eigenvalues are restricted to be smaller than a fixed barrier.

The paper is organized as follows. In section 2, we set up notations, we provide some mathematical preliminaries and we recall some known results for the Wishart random matrices as well as, for the sake of comparison, of Gaussian ensembles. Besides, this section also serves to set up our notations for the rest of the paper. In section 3, we outline the functional integration method followed by the steepest descent calculation. In subsection 3.1 we derive the left rate function explicitly for the special case $c = 1$ and in subsection 3.2 we extend the results to the case $c < 1$. In section 4 the analytical predictions are compared to numerical simulations. Section 5 concludes the paper with a summary and discussion, while the derivation of the rate function for $c < 1$ is given in the appendix.

2. Wishart and Gaussian random matrices: some preliminaries

We consider a rectangular $(M \times N)$ matrix X with M rows (representing M different samples) and N columns (representing N components of each sample). We assume that the entries of the matrix X are i.i.d random variables each drawn independently from a standard normal distribution, such that the joint distribution of the elements is given by $P[X] \propto \exp\left[-\frac{\beta}{2}\mathrm{tr}(X^\dagger X)\right]$ where the Dyson index $\beta = 1, 2$ corresponds respectively to real and complex matrices [16]. One then constructs the Wishart matrix $W = X^\dagger X$ by taking the product. The first natural question is: given the distribution of X, what is the joint distribution of the elements of W? It turns out that this is not quite easy to compute. For the case when $M \geqslant N$ (when the number of samples is larger or equal to the dimension of the vector), this was computed by Wishart [4]. The corresponding calculation for the opposite 'anti-Wishart' case, when $M < N$, turns out to be much more complicated. This was first obtained numerically [13] and only recently an analytical expression has been found [14, 15].

In contrast to the probability distribution of the matrix elements of W itself, the joint probability distribution (jpd) of its eigenvalues was known since a long time [29], and from it all the interesting spectral properties of the ensemble can be derived. We summarize them here together with the corresponding ones for the Gaussian ensemble.

2.1. Wishart (anti-Wishart) ensemble

For the case when $M \geqslant N$, all the N eigenvalues are positive and their jpd is given by

$$P_N(\lambda_1, \ldots, \lambda_N) = K_N \, \mathrm{e}^{-\frac{\beta}{2}\sum_{i=1}^N \lambda_i} \prod_{i=1}^N \lambda_i^{\frac{\beta}{2}(1+M-N)-1} \prod_{j<k} |\lambda_j - \lambda_k|^\beta \tag{7}$$

where K_N is a normalization constant and the parameter $\beta = 1, 2$ corresponds respectively to the real and complex X. On the other hand, for the anti-Wishart case $(M < N)$, there are only M positive eigenvalues (the rest of the $N - M$ eigenvalues are exactly 0) and their jpd is given exactly by the same formula as in (7) except that N and M are interchanged [15].

For the Wishart matrices with $M \geqslant N$, in the large N limit, the average density of states has the scaling form, $\rho_N(\lambda) \approx \frac{1}{N} f\left(\frac{\lambda}{N}\right)$ where $f(x)$ is the Marčenko–Pastur [17] function defined in (1). The corresponding result for the anti-Wishart case $(M < N)$, where one has M eigenvalues, simply follows by exchanging M and N.

For large N the maximum eigenvalue fluctuates around its average $\langle\lambda_{\max}\rangle \approx x_+(c)N$, and the typical fluctuation occurs over a scale of width $O(N^{1/3})$ around the mean. Johansson [12] and independently Johnstone [5] computed the limiting distribution of these *typical* fluctuations around the mean. They showed that for large N and for $c \leqslant 1$ [5, 12]

$$\lambda_{\max} = \left(\frac{1}{\sqrt{c}} + 1\right)^2 N + c^{1/6}\left(\frac{1}{\sqrt{c}} + 1\right)^{4/3} N^{1/3}\chi, \tag{8}$$

where the random variable χ has an N independent limiting distribution $\mathrm{Prob}(\chi \leqslant x) = F_\beta(x)$, which is the well-known Tracy–Widom distribution (see below).

2.2. Gaussian ensemble

In the case of a random $(N \times N)$ Gaussian matrix [27, 30], the jpd of eigenvalues is given by

$$P_N(\lambda_1, \ldots, \lambda_N) = B_N \, \mathrm{e}^{-\frac{\beta}{2}\sum_{i=1}^N \lambda_i^2} \prod_{j<k} |\lambda_j - \lambda_k|^\beta, \tag{9}$$

where B_N normalizes the jpd and $\beta = 1, 2$ and 4 corresponds respectively to the GOE (Gaussian orthogonal ensemble), GUE (Gaussian unitary ensemble) and GSE (Gaussian symplectic ensemble).

The average density of states in the large N limit has the scaling form $\rho_N(\lambda) \approx \frac{1}{\sqrt{N}} f_{\rm sc}\left(\frac{\lambda}{\sqrt{N}}\right)$, where $f_{\rm sc}(x)$ is the famous Wigner semi-circular law: $f_{\rm sc}(x) = \sqrt{\frac{1}{\pi}[2 - x^2]}$ with compact support over $x \in [-\sqrt{2}, \sqrt{2}]$.

Furthermore, the analogous asymptotic form of $\lambda_{\max}$ is known to be [31]

$$\lambda_{\max} = \sqrt{2N} + \frac{N^{-1/6}}{\sqrt{2}}\chi, \tag{10}$$

where the random variable χ again has the limiting N independent distribution, $\text{Prob}[\chi \leqslant x] = F_\beta(x)$.

In this paper, the main interest is focused on the largest eigenvalue. In summary, both the scaled variables $\lambda_{\max}/N$ in the Wishart case and $\lambda_{\max}/\sqrt{N}$ in the Gaussian case typically fluctuate over a region of width $\sim O(N^{-2/3})$ around their mean, and these typical fluctuations are described by the Tracy–Widom law $F_\beta(x)$.

The function $F_\beta(x)$, computed as a solution of a nonlinear Painlevé differential equation [31], approaches 1 as $x \to \infty$ and decays rapidly to zero as $x \to -\infty$. For example, for $\beta = 2$, $F_2(x)$ has the following tails [31]:

$$\begin{aligned} F_2(x) &\to 1 - O\left(\exp[-4x^{3/2}/3]\right) \qquad \text{as} \quad x \to \infty \\ &\to \exp[-|x|^3/12] \qquad\quad \text{as} \quad x \to -\infty. \end{aligned} \tag{11}$$

The probability density function $f_\beta(x) = \mathrm{d}F_\beta/\mathrm{d}x$ thus has highly asymmetric tails.

It follows from (8) that in the Wishart case, the Tracy–Widom distribution describes the probability of *typical and small* fluctuations of $\lambda_{\max}$ over a narrow region of width $\sim O(N^{1/3})$ around the mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$, where $x_+(c) = \left(\frac{1}{\sqrt{c}} + 1\right)^2$.

As mentioned in the introduction, in this paper we are concerned not with the *typical small* fluctuations of $O(N^{1/3})$ around the mean, but rather with *atypical large* fluctuations of $O(N)$. Thus, we are interested in computing the probability of extremely rare events. In fact, the question about the large deviation of the largest eigenvalue was addressed before in [12] and it was proved by rigorous methods that for $c \leqslant 1$ the probability of *large* fluctuations to the left of the mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$ behaves for large N as in (3), but an explicit expression for the left rate function $\Phi_-(x; c)$ was missing so far. On the other hand, for *large* fluctuations to the right of the mean $\langle\lambda_{\max}\rangle \approx x_+(c)N$,

$$1 - \text{Prob}[\lambda_{\max} \leqslant t, N] \sim \exp\left[-\frac{\beta}{2} N \Phi_+\left(\frac{t - x_+(c)N}{N}; c\right)\right] \tag{12}$$

for $t \sim O(N) \geqslant x_+(c)N$ located outside the Marčenko–Pastur sea to its right, and $\Phi_+(x; c)$ is the right rate function that was obtained explicitly in [12].

The purpose of this paper is to provide an exact result for $\Phi_-(x; c)$ for all $c \leqslant 1$. For the $c > 1$ (anti-Wishart) case, the result holds with M and N interchanged. Let us summarize our main results. For the case $c = 1$, we give an explicit expression for the left rate function $\Phi_-(x; 1)$ as stated in (36). Subsequently, the results in (5) and (6) follow. For $c < 1$, the function $\Phi_-(x; c)$ has a rather long analytical expression which is derived in the appendix. However, the function can be easily evaluated using Mathematica® as illustrated in figure 5.

These results should be compared to the corresponding ones for the Gaussian case. For the Gaussian ensemble, the left large deviations follow a similar law, namely,

$$\text{Prob}[\lambda_{\max} \leqslant t, N] \sim \exp\left[-\beta N^2 \Phi_-^{\rm Gauss}\left(\frac{\sqrt{2N} - t}{\sqrt{N}}\right)\right], \tag{13}$$

where $t \sim O(N^{1/2}) \leqslant \sqrt{2N}$ is located deep inside the Wigner sea. For the Gaussian case, $\langle\lambda\rangle = 0$. Thus, the corresponding question about the probability that all eigenvalues are less than their average is the same as the probability that all eigenvalues are negative. This probability plays a very important role in determining the average number of maxima of a random smooth potential, where a stationary point is a local maximum if all the eigenvalues of the associated Hessian matrix are negative. The calculation of this probability has been a subject of many theoretical and numerical studies with important applications in disordered systems, supercooled liquids, glassy models [32, 33] and, more recently, in anthropic principle-based string theory [34–36]. Very recently, the left rate function $\Phi_{-}^{\mathrm{Gauss}}(y)$ has been computed exactly using the functional integration methods [24]. Using this result, it was shown in [24] that for Gaussian matrices,

$$\mathrm{Prob}[\lambda_{\max} \leqslant 0; N] \sim \exp(-\theta N^2), \tag{14}$$

for large N where the coefficient

$$\theta = \frac{\beta}{4}\log 3 = 0.274\,653\ldots\beta. \tag{15}$$

In this paper, we adapt the techniques used in [24] for Gaussian matrices to the Wishart case. Similar techniques have recently been used also in other problems such as in the calculation of the average number of stationary points for Gaussian random field with N components in the large N limit [37, 38].

Incidentally, let us remark that our problem might also be tackled from the completely different viewpoint of zero-dimensional replica field theories, thanks to their recently discovered exact integrability [39]. This yet unexplored route may provide an independent derivation of our results.

3. Functional integration and the method of steepest descent

Our starting point is the joint distribution of eigenvalues of the Wishart ensemble in (7). Let $P_N(t)$ be the probability that the maximum eigenvalue $\lambda_{\max}$ is less than or equal to t. Clearly, this is also the probability that all the eigenvalues are less than or equal to t and can be expressed as a ratio of two multiple integrals

$$P_N(t) = \mathrm{Prob}[\lambda_{\max} \leqslant t] = \frac{Z_1(t)}{Z_0} = \frac{\int_0^t \ldots \int_0^t \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_N \exp\left(-\frac{\beta}{2}F[\vec{\lambda}]\right)}{\int_0^\infty \ldots \int_0^\infty \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_N \exp\left(-\frac{\beta}{2}F[\vec{\lambda}]\right)}, \tag{16}$$

where

$$F[\vec{\lambda}] = \sum_{i=1}^{N}\lambda_i - \left(1 + M - N - \frac{2}{\beta}\right)\sum_{i=1}^{N}\log\lambda_i - \sum_{j\neq k}\log|\lambda_j - \lambda_k|. \tag{17}$$

Written in this form, F mimics the free energy of a 2D Coulomb gas of interacting particles confined to the positive half-line ($\lambda > 0$) and subject to an external linear+logarithmic potential, as mentioned in the introduction. The denominator in (16), which is simply a normalization constant, represents the partition function of a free or 'unconstrained' Coulomb gas over $\lambda \in [0, \infty)$. The numerator, on the other hand, represents the partition function of the same Coulomb gas, but with the additional constraint that the gas is confined inside the box $\lambda \in [0, t]$, i.e. there is an additional wall or infinite barrier at the position $\lambda = t$. We will refer to the numerator as the partition function of a 'constrained' Coulomb gas. The constraint should not be effective when $t < x_-$ or $t > x_+$.

Note that in the Gaussian case, the external potential is harmonic over the whole real line ($V(\lambda) = \lambda^2/2$), while in the Wishart case, $V(\lambda) = \infty$ for $\lambda < 0$ (infinite barrier at $\lambda = 0$) and

$V(\lambda) = \lambda - (1 + M - N - 2/\beta)\log\lambda$ for $\lambda > 0$ representing a linear+logarithmic potential. By comparing the external potential and the logarithmic interaction term, it is easy to see that while for Gaussian ensembles a typical eigenvalue scales as $\lambda \sim \sqrt{N}$ for large N, for the Wishart case it scales as $\lambda \sim N$.

After defining the *constrained charge density*,

$$\hat{\varrho}_N(\lambda) := \varrho_N(\lambda; t) = \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda - \lambda_i)\theta(t - \lambda) \tag{18}$$

and taking into account the following trivial identity for a generic function $h(x)$:

$$\sum_{i=1}^{N} h(\lambda_i) = N\int \mathrm{d}\lambda\, \hat{\varrho}_N(\lambda)h(\lambda), \tag{19}$$

we may express, for large N, the partition function $Z_1(t)$ in (16) as a functional integral [24]:

$$Z_1(t) \propto \int \mathcal{D}[\hat{\varrho}_N]\exp\left\{-\frac{\beta}{2}\left[N\int_0^t \hat{\varrho}_N(\lambda)\lambda\,\mathrm{d}\lambda - N(M - N + 1 - 2/\beta)\int_0^t \hat{\varrho}_N(\lambda)\log\lambda\,\mathrm{d}\lambda \right.\right.$$
$$\left.\left. - N^2\int_0^t\int_0^t \hat{\varrho}_N(\lambda)\hat{\varrho}_N(\lambda')\log|\lambda - \lambda'|\,\mathrm{d}\lambda\,\mathrm{d}\lambda' - N\int_0^t \hat{\varrho}_N(\lambda)\log[\hat{\varrho}_N(\lambda)]\mathrm{d}\lambda\right]\right\}, \tag{20}$$

where the last entropic term is of order $O(N)$ and arises from the change of variables while going from an ordinary multiple integral to a functional integral, $[\{\lambda_i\}] \to [\hat{\varrho}_N(\lambda)]$. The constrained charge density $\hat{\varrho}_N(\lambda)$ satisfies the obvious constraints $\hat{\varrho}_N(\lambda) = 0$ for $\lambda > t$ and $\int_0^t \hat{\varrho}_N(\lambda)\,\mathrm{d}\lambda = 1$.

Since we are interested in fluctuations of $\sim O(N)$, it is convenient to work with the rescaled variables $\lambda = xN$ and $\zeta = t/N$. It is also reasonable to assume that for large N, the charge density scales accordingly as $\hat{\varrho}_N(\lambda) = N^{-1}\hat{f}(\lambda/N)$ so that $\hat{f}(x) = 0$ for $x > \zeta$ and $\int_0^\zeta \hat{f}(x)\,\mathrm{d}x = 1$.

In terms of the rescaled variables, the energy term in (20) becomes proportional to N^2 while the entropy term ($\sim O(N)$) is subdominant in the large N limit. Eventually, we can write

$$Z_1(\zeta) \propto \int \mathcal{D}[\hat{f}]\exp\left(-\frac{\beta}{2}N^2 S[\hat{f}(x); \zeta] + O(N)\right), \tag{21}$$

where

$$S[\hat{f}(x); \zeta] = \int_0^\zeta x\hat{f}(x)\,\mathrm{d}x - \alpha\int_0^\zeta \hat{f}(x)\log(x)\,\mathrm{d}x - \int_0^\zeta\int_0^\zeta \hat{f}(x)\hat{f}(x')\log|x - x'|\,\mathrm{d}x\,\mathrm{d}x'$$
$$+ C_1\left[\int_0^\zeta \hat{f}(x)\,\mathrm{d}x - 1\right], \tag{22}$$

where we have introduced the parameter $\alpha = \frac{1-c}{c}$ for later convenience. In (22), C_1 is a Lagrange multiplier enforcing the normalization of $\hat{f}$.

For large N we can evaluate the leading contribution to the action (22) by the method of steepest descent. This gives

$$Z_1(\zeta) \propto \exp\left[-\frac{\beta}{2}N^2 S[\hat{f}^\star(x); \zeta] + O(N)\right], \tag{23}$$

where $\hat{f}^\star$ is the solution of the stationarity condition

$$\frac{\delta S[\hat{f}(x); \zeta]}{\delta\hat{f}(x)} = 0. \tag{24}$$

Table 1. Values of L_1 and A in the different regions of the (c, ζ) plane.

	$c = 1$	$0 < c < 1$
$0 < \zeta < x_+$ (barrier *effective*)	$L_1 = 0$ (32)	L_1: see (50)
	$A = \frac{\zeta+4}{2}$ (32)	$A = \alpha\sqrt{\frac{\zeta}{L_1}} > \zeta$ (44)
$\zeta \geqslant x_+$ (barrier *ineffective*)	$L_1 = 0$	$L_1 = x_-$
	$A = \zeta = 4$	$A = \zeta = x_+$

This gives for $0 \leqslant x \leqslant \zeta$,

$$x - \alpha \log x + C_1 = 2 \int_0^\zeta \hat{f}(x') \log|x - x'| \, \mathrm{d}x'. \tag{25}$$

Differentiating (25) once with respect to x gives

$$\frac{1}{2} - \frac{\alpha}{2x} = \mathcal{P} \int_0^\zeta \frac{\hat{f}(x')}{x - x'} \, \mathrm{d}x', \qquad 0 \leqslant x \leqslant \zeta, \tag{26}$$

where $\mathcal{P}$ denotes the Cauchy principal part.

Finding a solution to the integral equation (26) is the main technical task. The following two subsections are devoted to the solution of (26), first for the special case $c = 1$ and then for $0 < c < 1$. We remark that the solution of (26) gives the average density of eigenvalues in the limit of large N for an ensemble of Wishart matrices whose rescaled eigenvalues are restricted to be smaller than the barrier ζ. We will refer to $\hat{f}(x)$ as the *constrained* spectral density.

Before proceeding to the technical points, it may be informative to first summarize the results for the constrained spectral density $\hat{f}(x)$ in the general $0 < c \leqslant 1$ case. The most general form is

$$\hat{f}(x) = \frac{1}{2\pi} \frac{\sqrt{x - L_1(c, \zeta)}}{\sqrt{\zeta - x}} \left[\frac{A(c, \zeta) - x}{x} \right], \tag{27}$$

where L_1 is the lower edge of the spectrum and A is related to the mutual position of the barrier with respect to the lower edge. In table 1, we schematically anticipate the values for L_1 and A in the different regions of the (c, ζ) plane.

The support of $\hat{f}$ is

$$L_1(c, \zeta) \leqslant x \leqslant \min[\zeta, A(c, \zeta)]. \tag{28}$$

At the lower edge of the support $L_1(c, \zeta)$, the density vanishes unless $c = 1$, in which case it diverges as $\sim 1/\sqrt{x}$.

At the upper edge of the support, according to the value of the minimum (ζ or $A(c, \zeta)$) in (28) the density can respectively diverge as $\sim 1/\sqrt{\zeta - x}$ or vanish.

Note that the unconstrained Marčenko–Pastur law (1) is recovered from (27) when the barrier is ineffective, i.e. $\zeta \geqslant x_+$.

3.1. The $c = 1$ case

In this case, the support of the unconstrained spectral density is $(0, 4]$, and the Marčenko–Pastur law prescribes an inverse square root divergence at $x = 0$. Furthermore, the density vanishes at $x = 4$ (see figure (1)).

In the constrained case, the barrier at ζ is only effective when $0 \leqslant \zeta \leqslant 4$. When the barrier crosses the point $\zeta = 4$ from below, the density shifts back again to the unconstrained case.

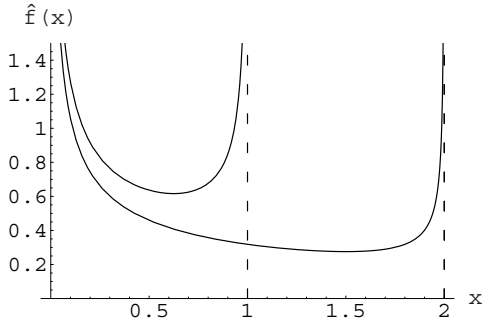

Figure 2. Constrained spectral density $\hat{f}(x)$ for the barrier at $\zeta = 1$ and $\zeta = 2$.

The integral equation for $\hat{f}(x)$ (26) becomes

$$\frac{1}{2} = \mathcal{P}\int_0^\zeta \frac{\hat{f}(x')}{x - x'}\,\mathrm{d}x', \qquad 0 \leqslant x \leqslant \zeta. \tag{29}$$

The general solution of equations of the type

$$\mathcal{P}\int_0^\zeta \frac{\hat{f}(x')}{x - x'}\,\mathrm{d}x' = g(x) \tag{30}$$

is given by Tricomi's theorem [40]:

$$\hat{f}(x) = \frac{1}{\pi^2\sqrt{x(\zeta - x)}}\left[\mathcal{P}\int_0^\zeta \sqrt{\omega(\zeta - \omega)}\,\frac{g(\omega)}{\omega - x}\,\mathrm{d}\omega + B\right], \tag{31}$$

where B is an arbitrary constant. After putting $g(\omega) = 1/2$ in (31) and determining B by the normalization condition $\int_0^\zeta \hat{f}(x)\,\mathrm{d}x = 1$, we finally get

$$\hat{f}(x) = \frac{1}{2\pi\sqrt{x(\zeta - x)}}\left[\frac{\zeta}{2} + 2 - x\right], \qquad 0 \leqslant x \leqslant \zeta. \tag{32}$$

A plot of this charge density for two values of the barrier ζ is given in figure 2. In summary, the average density of states with a barrier at ζ is given by

$$\hat{f}(x) = \begin{cases} \frac{1}{2\pi\sqrt{x(\zeta - x)}}\left[\frac{\zeta}{2} + 2 - x\right] & 0 \leqslant \zeta \leqslant 4 \\ \frac{1}{2\pi}\sqrt{\frac{4-x}{x}} & \zeta \geqslant 4. \end{cases} \tag{33}$$

Thus, for all $\zeta > 4$, the solution sticks to the $\zeta = 4$ case. Note that both cases in (33) can be obtained from the general formula (27).

Now we can substitute (33) back into (25) to find the value of the multiplier C_1 and eventually evaluate the action $S[f^\star(x); \zeta]$ (22) explicitly for $0 \leqslant \zeta \leqslant 4$:

$$S(\zeta) := S[\hat{f}^\star(x); \zeta] = 2\log 2 - \log\zeta + \frac{\zeta}{2} - \frac{\zeta^2}{32}. \tag{34}$$

From (23), we get $Z_1(\zeta) \approx \exp(-\beta N^2 S(\zeta)/2)$. For the denominator, $Z_0 = Z_1(\zeta = \infty) = Z_1(\zeta = 4) \approx \exp(-\beta N^2 S(4)/2)$, where we have used the fact that the solution for any $\zeta > 4$ (e.g., when $\zeta = \infty$) is the same as the solution for $\zeta = 4$. Thus, eventually the probability $P_N(t)$ (16) decays for large N as

$$P_N(t) = \frac{Z_1(t)}{Z_0} \approx \exp\left\{-\frac{\beta}{2}N^2[S(\zeta) - S(4)]\right\} \approx \exp\left\{-\frac{\beta}{2}N^2\Phi_-\left(\frac{4N - t}{N}; 1\right)\right\}, \tag{35}$$

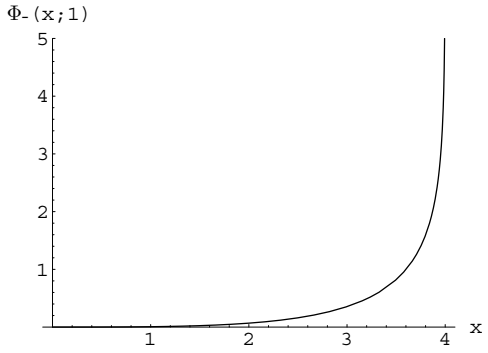

Figure 3. Rate function $\Phi_-(x;1)$ (see (36)).

where the rate function is given by

$$\Phi_-(x;1) = \begin{cases} 2\log 2 - \log(4-x) - \dfrac{x}{4} - \dfrac{x^2}{32} & x \geqslant 0 \\ 0 & x \leqslant 0 \end{cases} \tag{36}$$

and is plotted in figure 3.

We now turn to the original problem of determining the probability of the following extremely rare event, i.e. that all the eigenvalues happen to lie below the mean value $\langle\lambda\rangle = \int_0^{4N} \lambda\rho_N(\lambda)\,\mathrm{d}\lambda = N$. Starting from (35), this is easily computed by putting the barrier at the mean value $t = N$, i.e. $\zeta = 1$. We then get for large N

$$\mathrm{Prob}\,[\lambda_{\max} \leqslant \langle\lambda\rangle = N, N] \sim \exp[-\theta(1)N^2], \tag{37}$$

where

$$\begin{aligned}\theta(1) &= \frac{\beta}{2}\Phi_-(3;1) \\ &= \beta\left(\log 2 - \frac{33}{64}\right).\end{aligned} \tag{38}$$

Since we are calculating the probability of negative fluctuations of $O(N)$ to the left of the mean $\langle\lambda_{\max}\rangle = x_+(c)N$ of $\lambda_{\max}$, when the argument of the left rate function $\Phi_-(x;1)$ is very small (i.e. for fluctuations $\ll O(N)$), (35) should smoothly match with the left tail of the Tracy–Widom distribution that describes fluctuations of $\sim O(N^{1/3})$ to the left of the mean. Indeed, from (36), as $x \to 0$:

$$\Phi_-(x;1) \approx \frac{x^3}{192}. \tag{39}$$

Substituting this small x behavior in (35) we get, for fluctuations $\ll O(N)$ to the left of the mean,

$$\begin{aligned}P_N(t) &\sim \exp\left[-\frac{\beta}{384}N^2(4-t/N)^3\right] \\ &= \exp[-|\chi|^3/12],\end{aligned} \tag{40}$$

where $\chi = (t-4N)/(2^{4/3}N^{1/3})$. This coincides with Johansson's result for the Tracy–Widom fluctuations in (8) for $c = 1$ and comparing (40) and (11), we see that we recover the left tail of the Tracy–Widom distribution.

3.2. The $c < 1$ case

Our approach is very similar to the previous case. However, some additional technical subtleties, which we will emphasize, arise in this case.

As in the unconstrained case, we expect a lower bound $L_1 \equiv L_1(c, \zeta)$ to the support of the constrained $\hat{f}(x)$. The parameter L_1 will be determined later through the normalization condition for $\hat{f}(x)$.

It is convenient to reformulate (26) in terms of the new variable $y = x - L_1$, measuring the distance with respect to the lower edge of the support, where $\hat{f}(x)$ is assumed to vanish.

Equation (26) then reads

$$\frac{1}{2} - \frac{\alpha}{2(y + L_1)} = \mathcal{P} \int_0^L \frac{\tilde{f}(y')}{y - y'} \, \mathrm{d}y' \qquad 0 \leqslant y \leqslant L, \tag{41}$$

where we have denoted $L = \zeta - L_1$ and $\tilde{f}(y) \equiv \hat{f}(y + L_1)$.

The general solution of (41) in this case is

$$\tilde{f}(y) = \frac{1}{\pi \sqrt{y(L - y)}} \left[-\frac{y}{2} - \frac{\alpha}{2} \frac{\sqrt{L_1(L + L_1)}}{y + L_1} + B' \right], \tag{42}$$

and the constant B' is determined by the condition $\tilde{f}(y = 0) = 0$. Thus, we get

$$\tilde{f}(y) = \frac{\sqrt{y}}{2\pi \sqrt{L - y}} \left[\frac{A - L_1 - y}{y + L_1} \right], \tag{43}$$

where

$$A \equiv A(c, \zeta) = \alpha \sqrt{\zeta / L_1}. \tag{44}$$

Note that everything is expressed in terms of the only still unknown parameter L_1.

From (43) we can infer two kinds of possible behaviours for $\tilde{f}(y)$ due to the competing effects of the singularity for $y \to L$ (where the denominator vanishes) and the suppression for $y \to A - L_1$ (where the numerator vanishes).

Thus, depending on which of the following two conditions applies once we have put the barrier at ζ:

$$\begin{aligned} &A(c, \zeta) - L_1(c, \zeta) > L(c, \zeta) \to \sqrt{L_1(c, \zeta)\zeta} < \alpha \quad \text{(I)} \\ &A(c, \zeta) - L_1(c, \zeta) < L(c, \zeta) \to \sqrt{L_1(c, \zeta)\zeta} > \alpha \quad \text{(II)}, \end{aligned} \tag{45}$$

$\tilde{f}$ can diverge at $y = L$ or vanish at $A - L_1$ respectively. In (45) we have restored the functional dependence for clarity.

This is a subtle point because, given the barrier at ζ, we cannot determine *a priori* which of the previous conditions holds. In fact, $L_1(c, \zeta)$ should be determined *a posteriori* separately for each case from the normalization condition

$$\int_0^L \tilde{f}(y) \, \mathrm{d}y = 1. \tag{46}$$

Once this is done, it turns out that conditions (45) can be reformulated in terms of the position of the barrier ζ in the following much simpler way:

$$\begin{aligned} &0 < \zeta < x_+ \qquad \text{(I)} \\ &\zeta \geqslant x_+ \qquad \text{(II)}. \end{aligned} \tag{47}$$

We summarize here the final results in the two cases.

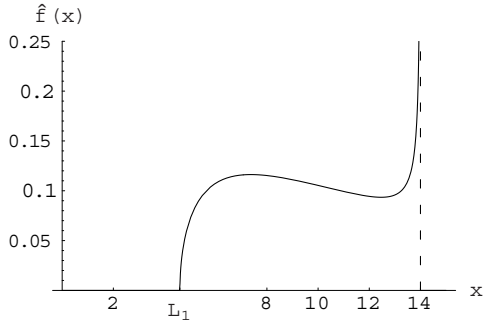

Figure 4. Constrained spectral density $\hat{f}(x)$ for $c = 0.1$ and $\zeta = 14$.

3.2.1. Case I. $0 < \zeta < x_+$. The normalization condition (46) leads to the following cubic equation for $w \equiv w(c, \zeta) = \sqrt{L_1(c, \zeta)}$:

$$w^3 - [2(2+\alpha) + \zeta]w + 2\alpha\sqrt{\zeta} = 0, \tag{48}$$

which has always three real solutions, one negative (w_0) and two positive,

$$w_k(c, \zeta) = \frac{2p}{3\varrho^{1/3}} \cos\left(\frac{\theta + 2k\pi}{3}\right), \qquad k = 0, 1, 2, \tag{49}$$

where

$$\begin{cases} p = -[2(2+\alpha) + \zeta] \\ q = 2\alpha\sqrt{\zeta} \\ B = -\left(\dfrac{q^2}{4} + \dfrac{p^3}{27}\right) \\ \varrho = \sqrt{-p^3/27} \\ \theta = \arctan\left(\dfrac{2\sqrt{B}}{q}\right). \end{cases}$$

Note that $w_2 < w_1$. With simple considerations, we conclude that the right root to be chosen is $w_2(c, \zeta)$. Thus,

$$L_1(c, \zeta) = w_2^2(c, \zeta). \tag{50}$$

Finally, we can write down the full constrained unshifted spectral density as

$$\hat{f}(x) = \frac{1}{2\pi} \frac{\sqrt{x - L_1(c, \zeta)}}{\sqrt{\zeta - x}} \left[\frac{A(c, \zeta) - x}{x}\right], \tag{51}$$

valid for $L_1(c, \zeta) \leqslant x \leqslant \zeta$ where $L_1(c, \zeta)$ is given by (50) and $A(c, \zeta)$ by (44).

A plot of $\hat{f}(x)$ for $c = 0.1$ and $\zeta = 14$ is given in figure 4. In this case, $L_1(c, \zeta) \approx 4.600\,84$ and $A(c, \zeta) \approx 15.6996$.

3.2.2. Case II. $\zeta \geqslant x_+$. In this case, the barrier is immaterial and we should recover the unconstrained Marčenko–Pastur distribution. The support of $\tilde{f}(y)$ is $[0, A - L_1]$ and this implies that we can safely put $L = A - L_1$ in (46).

The integration can be performed and coming back to the unshifted spectral density $\hat{f}(x)$, we get

$$\hat{f}(x) = \frac{1}{2\pi} \frac{\sqrt{x - L_1}\sqrt{L_2 - x}}{x}, \tag{52}$$

Table 2. Some values of the rate function (see the text for further explanation).

c	$\Phi_-\left(\frac{2}{\sqrt{c}}+1;c\right)$
0.1	0.475 802
0.2	0.449 162
0.4	0.414 592
0.6	0.390 245
0.8	0.371 04
0.95	0.358 805
1	0.355 044

valid for $L_1 \leqslant x \leqslant L_2$, where

$$\begin{cases} L_1 = x_- \\ L_2 = L_1 + L = x_+, \end{cases} \tag{53}$$

which is the unconstrained Marčenko–Pastur distribution, as it should.

Obviously, the interesting case for computing large fluctuations is case I, but before that it is interesting to evaluate the limit $c \to 1^-$ in (51) and (52) in order to recover the result (33) in subsection 3.1.

The case of equation (52) is straightforwardly obvious. For the other, it is a matter of simple algebra to show that

$$\lim_{c\to 1^-} L_1(c,\zeta) = 0 \tag{54}$$

$$\lim_{c\to 1^-} A(c,\zeta) = (\zeta+4)/2, \tag{55}$$

so that (51) perfectly matches (33).

Furthermore, cases I and II should match smoothly as ζ hits precisely the limiting value x_+. This corresponds to $A(c,\zeta) \equiv \zeta \to A(c,x_+) \equiv x_+$. It is again straightforward to check that this last condition implies $L_1(c,x_+) \equiv x_-$ so that (51) recovers (52).

Coming back to the large fluctuation problems for case I, we would like to insert (51) into (22) in order to evaluate (23). It turns out that the integrals involved can be analytically solved in terms of derivatives of hypergeometric functions, but a more explicit formula is derived in the appendix. We will give here a plot of the rate function $\Phi_-(x;c)$ that describes the large fluctuations of $O(N)$ to the left of the mean $\langle\lambda_{\max}\rangle = x_+(c)N$:

$$\begin{aligned} P_N(t) = \frac{Z_1(t)}{Z_0} &\approx \exp\left\{-\frac{\beta}{2}N^2[S(\zeta)-S(x_+)]\right\} \\ &= \exp\left\{-\frac{\beta}{2}N^2\Phi_-\left(x_+ - \frac{t}{N};c\right)\right\} \end{aligned} \tag{56}$$

The plot is given in figure 5 for several values of c approaching 1. The limiting case $\Phi_-(x;1)$ (36) is also plotted.

We can now compute to the leading order the probability that all the eigenvalues are less than the mean value $\langle\lambda\rangle = N/c$. This amounts to putting the barrier at $t = N/c$ in (56), which gives $\Phi_-\left(\frac{2}{\sqrt{c}}+1;c\right)$. Several numerical values are given in table 2.

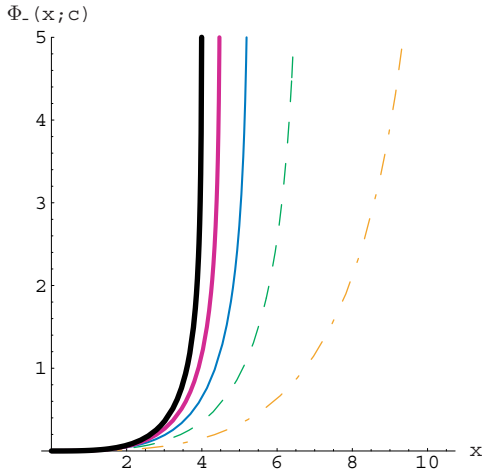

Figure 5. Rate function $\Phi_-(x;c)$ for the following values (from right to left) of $c = 0.2, 0.4, 0.6, 0.8, 1$. See also figure 3.

4. Numerical results

Formulae (33), (35), (51) and (56) have been numerically checked on samples of Hermitian matrices ($\beta = 2$) up to $N = 30$, $M = 300$ and the agreement with the analytical results is already excellent. We describe in this section the numerical methods and results.

A direct sampling of Wishart matrices up to those sizes is computationally very demanding. We applied the following much faster technique, suggested in [22].

Let $L_\beta = B_\beta B_\beta^T$ be the tridiagonal matrix corresponding to

$$B_\beta \sim \begin{pmatrix} \chi_{2a} & & & \\ \chi_{\beta(N-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_\beta & \chi_{2a-\beta(N-1)} \end{pmatrix}. \tag{57}$$

B_β is a square $N \times N$ matrix with nonzero entries on the diagonal and subdiagonal and $a = (\beta/2)M$. The nonzero entries χ_k are independent random variables obtained from the square root of a χ^2-distributed variable with k degrees of freedom. It has been proved in [22] that L_β has the same joint probability distribution of eigenvalues as (7). Thus, as far as we are interested in eigenvalue properties, we can use the L_β ensemble instead of the original Wishart one. This makes the diagonalization process much faster due to the tridiagonal structure of the matrices L_β.

We report the following four plots: the first two (figures 6 and 7) are for the case $c = 1$ and the last two (figures 8 and 9) for the case $c = 0.1$.

In figure 6, we plot the histogram of normalized eigenvalues $\lambda/2N$ for an initial sample of 3×10^5 Hermitian matrices ($\beta = 2$, $N = M = 30$), such that the matrices with $\lambda_{\max}/2N > \zeta$ are discarded. The barrier is located at $\zeta = 3$. On top of it, we plot the theoretical distribution (33).

In figure 8, we do the same but in the case $N = 10$, $M = 100$. The barrier is located at $\zeta = 14$. The theoretical distribution is now taken from (51).

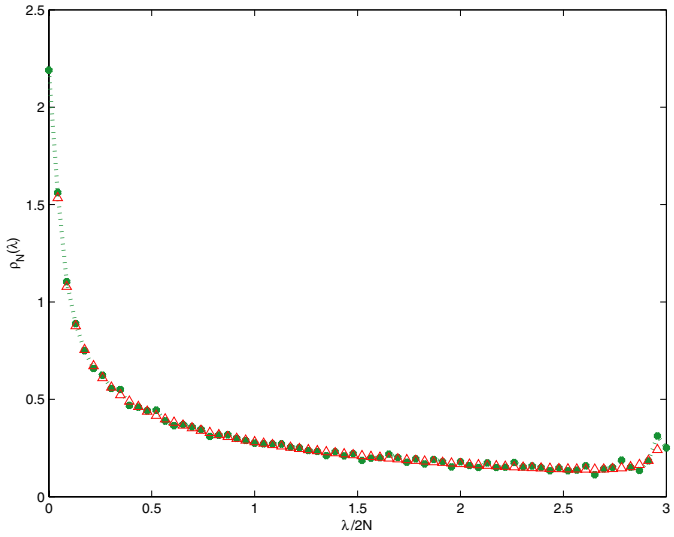

Figure 6. Constrained spectral density $\hat{\varrho}_N(\lambda)$ for $N = M = 30$. The barrier is at $\zeta = 3$. In dotted green is the histogram of rescaled eigenvalues over an initial sample of 3×10^5 matrices ($\beta = 2$). In red (triangles) is the theoretical distribution.

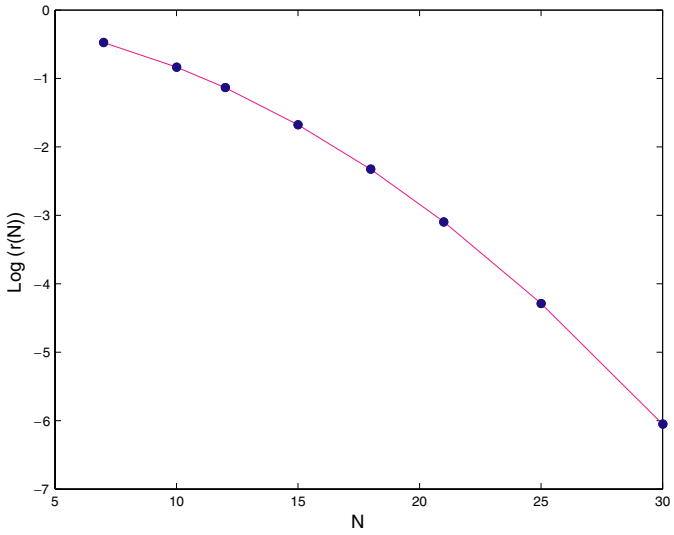

Figure 7. Natural logarithm of the probability that all the rescaled eigenvalues are less than $\zeta = 3$ versus N for the case $c = 1$ ($x_+ = 4$). The data points are fitted with a parabola (solid line).

To obtain the plots in figures 7 and 9, we generate $\approx 5 \times 10^5 L_2$ matrices for different values of $N = 7 \to 30$ (or 15). The parameters (c, ζ) are kept fixed to the value $(1, 3)$ for figure 7 ($x_+ = 4$) and $(0.1, 14)$ for figure 9 ($x_+ \approx 17.32$). The constraining capability of those barriers can be evaluated by the ratio $\kappa(\zeta, c) = (x_+ - \zeta)/(x_+ - x_-)$, corresponding to the window of forbidden values for the largest eigenvalue. We get $\kappa(3, 1) = 0.25$ and $\kappa(14, 0.1) \approx 0.26$, to be compared with the values of $\kappa(\zeta, c) = (2 + \sqrt{c})/4$ for the barrier at the mean value

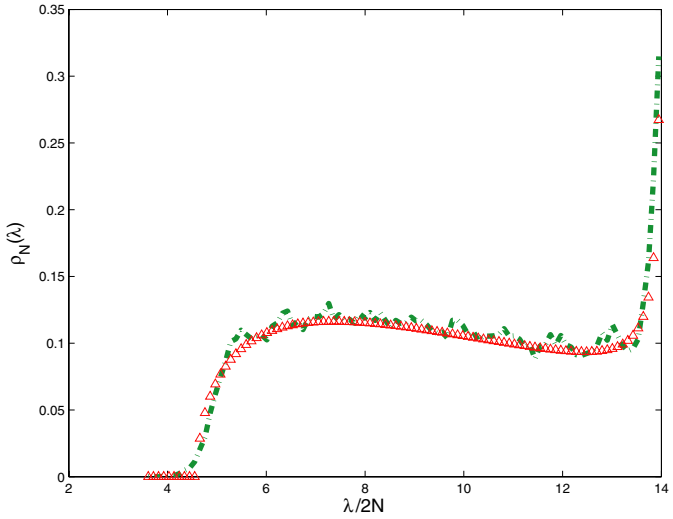

Figure 8. Constrained spectral density $\hat{\varrho}_N(\lambda)$ for $N = 10$, $M = 100$ ($c = 0.1$). The barrier is at $\zeta = 14$. In green (dash-dot) is the histogram of rescaled eigenvalues over an initial sample of 5×10^5 matrices ($\beta = 2$). In red (triangles) is the theoretical distribution.

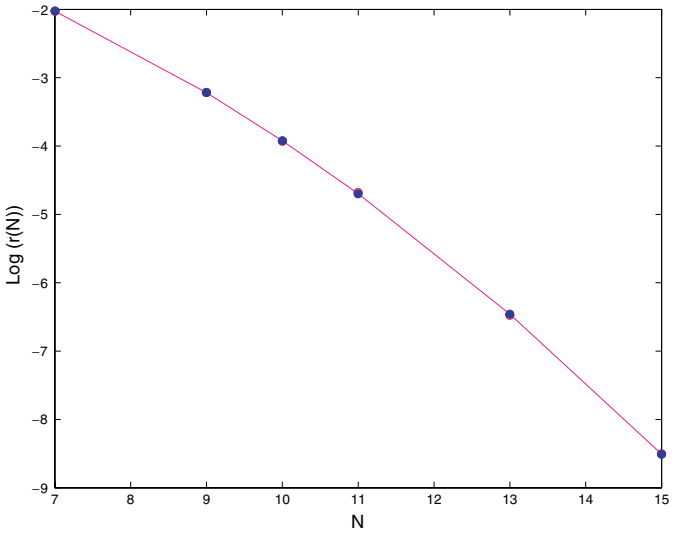

Figure 9. Natural logarithm of the probability that all the rescaled eigenvalues are less than $\zeta = 14$ versus N for the case $c = 0.1$ ($x_+ \approx 17.32$). The data points are fitted with a parabola (solid line).

$\zeta = 1/c$, which would respectively give $\kappa = 0.75$ and $\kappa \approx 0.58$. This relative mildness of the constraint allows us to get a much more reliable and faster statistics in the simulations.

For each value of N, we determine the empirical frequency $r(N)$ of constrained matrices as the ratio of the number of matrices whose largest rescaled eigenvalue is less than ζ to the total number of samples (5×10^5). The logarithm of $r(N)$ versus the size N is then naturally fitted by a parabola $aN^2 + bN + \hat{c}$ to test the prediction for a in formulae (35) and (56).

The best values for the coefficient a of the leading term are estimated as $-0.006\,153$ ($c = 1$) and -0.0357 ($c = 0.1$), to be compared respectively with the theoretical prediction $\Phi_-(1;1) \approx -0.006\,432$ and $\Phi_-(x_+ - 14; 0.1) \approx -0.036\,66$. Despite the relatively small sizes and the $O(N)$ corrections, the agreement is already good.

5. Conclusions

In this paper we have studied the probability of atypically large negative fluctuations (with respect to the mean) of the largest eigenvalue $\lambda_{\max}$ of a random Wishart matrix. The standard Coulomb gas analogy for the joint probability distribution of eigenvalues allowed us to use the tools of statistical physics, such as the functional integral method evaluated for large N by the method of steepest descent. Using these tools, we have analytically computed the probability of large deviations of $\lambda_{\max}$ to the left of its mean. In particular, our main motivation was to compute the probability of a rare event: all eigenvalues are less than the average $\langle\lambda\rangle = N/c$. This implies that the largest eigenvalue itself is less than $\langle\lambda\rangle = N/c$. This question is relevant in estimating the efficiency of the 'principal components analysis' method used in multivariate statistical analysis of data. Our main result is to show that, to leading order in N, this probability decays as $\sim\exp\left[-\frac{\beta}{2}N^2\Phi_-\left(\frac{2}{\sqrt{c}}+1;c\right)\right]$, where $\Phi_-(x;c)$ is a rate function that we have explicitly computed. The quadratic, instead of linear, N dependence of the exponential reflects the eigenvalue correlations.

Furthermore, our method allows us to determine exactly the functional form of the constrained spectral density, i.e. the average charge density of a Coulomb gas constrained to be within a finite box $\lambda \in [0, t]$.

All the analytical results are in excellent agreement with the numerical simulations on samples of Hermitian matrices up to $N = 30$, and the estimates of the large deviation prefactor are already good even for $N \sim 15$.

Acknowledgments

One of us (PV) has been supported by a Marie Curie Early Stage Training Fellowship (NET-ACE project). We are grateful to Gernot Akemann, Igor Krasovsky and Yang Chen for helpful comments and for pointing out to us relevant references. The support by Sergio Consoli (Brunel) for the numerical simulations is also gratefully acknowledged.

Appendix. Rate function for $c < 1$

We evaluate in a closed form the action $S(\zeta) := S[\hat{f}^\star(x); \zeta]$ (see (22)) for the case $c < 1$, where $\hat{f}^\star(x)$ is given by (51). The result is equation (A.3).

The rate function $\Phi_-(x; c)$ for $c < 1$, given by

$$\Phi_-(x;c) = S(x_+ - x) - S(x_+) \tag{A.1}$$

can be evaluated immediately.

After inserting (51) into (22) and determining C_1 from (25), we find that $S(\zeta)$ is given by

$$S(\zeta) = \frac{1}{2}\int_{L_1}^{\zeta} \hat{f}(x)x\,\mathrm{d}x - \frac{\alpha}{2}\int_{L_1}^{\zeta} \hat{f}(x)\log(x)\,\mathrm{d}x - \int_{L_1}^{\zeta} \hat{f}(x)\log(x - L_1)\,\mathrm{d}x + \frac{L_1}{2} - \frac{\alpha}{2}\log(L_1). \tag{A.2}$$

After the substitution $x = (\zeta - L_1)t + L_1$ in the integrals in (A.2) and some simple algebra, $S(\zeta)$ can be expressed as

$$S(\zeta) = -\frac{\alpha}{2}\Theta_1 - \Theta_2 + \frac{\zeta - L_1}{2}\Xi + \frac{L_1}{2} - \frac{\alpha}{2}\log(L_1), \tag{A.3}$$

where Θ_k and Ξ are the following functions of c and ζ:

$$\begin{aligned}\Theta_k &= \frac{\zeta - L_1}{2\pi}\left\{\log(\zeta - L_1)\left[\frac{A}{\zeta - L_1}\mathcal{I}_0\left(\frac{L_1}{\zeta - L_1}\right) - \frac{\pi}{2}\right] + \frac{A}{\zeta - L_1}\mathcal{I}_k\left(\frac{L_1}{\zeta - L_1}\right)\right\} \\ \Xi &= \frac{A}{4} - \frac{3}{16}\zeta - \frac{L_1}{16} + \frac{\alpha}{2\pi}\mathcal{I}_3\left(\frac{L_1}{\zeta - L_1}\right) + \frac{1}{2} - \log(2).\end{aligned} \tag{A.4}$$

The functions $\mathcal{I}_k(x)$ are given by the following integrals, which are then computed explicitly in a closed form:

$$\mathcal{I}_0(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathcal{I}_3(x) \tag{A.5}$$

$$\mathcal{I}_1(x) = \int_0^1 \mathrm{d}t \frac{\log(t + x)}{t + x}\sqrt{\frac{t}{1 - t}} \tag{A.6}$$

$$\mathcal{I}_2(x) = \int_0^1 \mathrm{d}t \frac{\log t}{t + x}\sqrt{\frac{t}{1 - t}} \tag{A.7}$$

$$\mathcal{I}_3(x) = \int_0^1 \mathrm{d}t \log(t + x)\sqrt{\frac{t}{1 - t}}. \tag{A.8}$$

We proceed to compute the three integrals in (A.6)–(A.8) following very closely [28, paper I, appendix B].

The integral $\mathcal{I}_3(x)$ (and thus also $\mathcal{I}_0(x)$) can be computed by Mathematica®:

$$\mathcal{I}_3(x) = \frac{\pi}{2}\left[1 + 2x - 2\sqrt{x(1 + x)} + 2\log\left[1 + \sqrt{1 + \frac{1}{x}}\right] + \log\left(\frac{x}{4}\right)\right], \tag{A.9}$$

while $\mathcal{I}_1(x)$ and $\mathcal{I}_2(x)$ are given in terms of derivatives of hypergeometric functions. More explicit expressions can be given as follows, starting with $\mathcal{I}_1(x)$. Exploiting the identity $h^\lambda \log h = \partial_\lambda h^\lambda$, we can rewrite the integral as

$$\mathcal{I}_1(x) = \left[\partial_\lambda \int_0^1 \mathrm{d}t\,(t + x)^\lambda \sqrt{\frac{t}{1 - t}}\right]\Bigg|_{\lambda=-1}, \tag{A.10}$$

and the integral in (A.10) can be evaluated in terms of Kummer's hypergeometric function:

$$\mathcal{I}_1(x) = \frac{\pi}{2}\left\{\partial_\lambda\left[x^\lambda\, {}_2F_1\left(-\lambda, \frac{3}{2}; 2; -\frac{1}{x}\right)\right]\right\}\Bigg|_{\lambda=-1}. \tag{A.11}$$

Now, applying the transformation formulae [41, 15.3.7, page 559] and evaluating the derivatives of Gamma functions that arise, we finally get

$$\mathcal{I}_1(x) = \frac{\pi}{2}\left[-2\log 4 + 2\hat{i}_1(x) - 2\sqrt{\frac{x}{1 + x}}\log\left(\frac{4x}{e^2}\right) - 2\sqrt{x}\hat{i}_2(x)\right], \tag{A.12}$$

where

$$\hat{i}_1(x) = [\partial_\mu\, {}_2F_1(1 - \mu, -\mu; -\mu + 1/2; -x)]|_{\mu=0} \tag{A.13}$$

$$\hat{i}_2(x) = \left[\partial_\mu (1 + x)^{\mu - 1/2}\, {}_2F_1(\mu, \mu + 1; \mu + 3/2; -x)\right]\Big|_{\mu=0}. \tag{A.14}$$

To evaluate $\hat{i}_1(x)$ and $\hat{i}_2(x)$, we use the series expansion for hypergeometric functions [41, 15.1.1, page 556] and upon differentiation we get

$$\hat{i}_1(x) = -\sum_{n=1}^{\infty} B\left(\frac{1}{2}, n\right)(-x)^n, \tag{A.15}$$

where $B(v, w) = \frac{\Gamma(v)\Gamma(w)}{\Gamma(v+w)}$ is Euler's beta function. Introducing the integral representation of the beta function

$$B(x, y) = \int_0^1 \mathrm{d}t\, t^{x-1}(1-t)^{y-1} \tag{A.16}$$

into (A.15) and upon exchanging summation and integral, we arrive with the help of $\sum_{n=0}^{\infty}(-xt)^n = (1+xt)^{-1}$ to

$$\hat{i}_1(x) = x\int_0^1 \frac{\mathrm{d}t}{\sqrt{1-t}(1+xt)} = 2\sqrt{\frac{x}{1+x}}\,\mathrm{arcsinh}(\sqrt{x}). \tag{A.17}$$

Following the same procedure, we get for $\hat{i}_2(x)$,

$$\hat{i}_2(x) = \frac{1}{\sqrt{1+x}}[\log(1+x) - i_1(x)], \tag{A.18}$$

where $i_1(x)$ is defined in [28] as

$$i_1(x) = -2 + 2\sqrt{\frac{1+x}{x}}\,\mathrm{arctanh}\left(\sqrt{\frac{x}{1+x}}\right). \tag{A.19}$$

From (A.12) we get the final result for $\mathcal{I}_1(x)$:

$$\mathcal{I}_1(x) = \pi\left\{-\log 4 + \sqrt{\frac{x}{1+x}}\left[2\mathrm{arcsinh}(\sqrt{x}) + 2\sqrt{1+\frac{1}{x}}\mathrm{arctanh}\left(\sqrt{\frac{x}{1+x}}\right) - \log[4x(1+x)]\right]\right\}. \tag{A.20}$$

Following the very same procedure as in the previous case, we find for $\mathcal{I}_2(x)$:

$$\mathcal{I}_2(x) = \pi\left[-\log 4 + \sqrt{\frac{x}{x+1}}(2\mathrm{arcsinh}(\sqrt{x}) - \log(x))\right]. \tag{A.21}$$

Now we compute the limit $c \to 1^-$ in (A.3) to recover (34). Given that, for $c \to 1^-$, $L_1 \to 0$, $\alpha \to 0$ and $A \to (\zeta+4)/2$, we have to evaluate the integrals $\mathcal{I}_k(x)$ for $x \to 0$. This gives

$$\mathcal{I}_0(0) \sim \pi \tag{A.22}$$

$$\mathcal{I}_1(0) \sim -\pi\log 4 \tag{A.23}$$

$$\mathcal{I}_2(0) \sim -\pi\log 4 \tag{A.24}$$

$$\mathcal{I}_3(0) \sim -\frac{\pi}{2}(\log 4 - 1). \tag{A.25}$$

Then, $S[\hat{f}^\star(x);\zeta]\big|_{c\to 1^-} \sim \left[-\Theta_2 + \frac{\zeta}{2}\Xi\right]\big|_{c\to 1^-} = -\log\zeta + 2\log 2 + \frac{\zeta}{2} - \frac{\zeta^2}{32}$ as it should (see equation (34)).

References

[1] Wilks S S 1962 *Mathematical Statistics* (New York: Wiley)
[2] Fukunaga K 1990 *Introduction to Statistical Pattern Recognition* (New York: Elsevier)
[3] For a nice pedagogical introduction to PCA, see Smith L I 2002 A tutorial on principal components analysis (http://csnet.otago.ac.nz/cosc453/student_tutorials/principal.components.pdf)
[4] Wishart J 1928 *Biometrica* **20** 32
[5] Johnstone I M 2001 *Ann. Stat.* **29** 295
[6] Preisendorfer R W 1988 *Principal Component Analysis in Meteorology and Oceanography* (New York: Elsevier)
[7] Bouchaud J-P and Potters M 2001 *Theory of Financial Risks* (Cambridge: Cambridge University Press)
[8] Burda Z and Jurkiewicz J 2004 *Physica* A **344** 67
Burda Z, Jurkiewicz J and Waclaw B 2005 *Acta Phys. Pol.* B **36** 2641 and references therein
[9] Telatar E 1999 *Eur. Trans. Telecommun.* **10** 585
[10] Fyodorov Y V and Sommers H-J 1997 *J. Math. Phys.* **38** 1918
Fyodorov Y V and Khoruzhenko B A 1999 *Phys. Rev. Lett.* **83** 66
[11] Shuryak E V and Verbaarschot J J M 1993 *Nucl. Phys.* A **560** 306
Verbaarschot J J M 1994 *Phys. Rev. Lett.* **72** 2531
[12] Johansson K 2000 *Commun. Math. Phys.* **209** 437
[13] Maslov S and Zhang Y C 2001 *Phys. Rev. Lett.* **87** 248701
[14] Yu Y K and Zhang Y C 2002 *Physica* A **312** 1
[15] Janik R A and Nowak M A 2003 *J. Phys. A: Math. Gen.* **36** 3629
[16] Dyson F J 1962 *J. Math. Phys.* **3** 140
[17] Marčenko V A and Pastur L A 1967 *Math. USSR-Sb.* **1** 457
[18] Dyson F J 1971 *Rev. Mex. Fis.* **20** 231
[19] Bohigas O and Flores J 1971 *Rev. Mex. Fis.* **20** 217
[20] Edelman A 1988 *SIAM J. Matrix Anal. Appl.* **9** 543
[21] Forrester P J 1993 *Nucl. Phys.* B **402** 709
[22] Dumitriu I and Edelman A 2002 *J. Math. Phys.* **43** 5830
[23] Edelman A and Sutton B D 2006 *Preprint* math-ph/0607038
[24] Dean D S and Majumdar S N 2006 *Phys. Rev. Lett.* **97** 160201
[25] Hiai F and Petz D 2000 *The semicircle law, free random variables and entropy* (Providence, RI: American Mathematical Society)
[26] Deift P, Its A and Krasovsky I 2006 *Preprint* math.FA/0609451
[27] Mehta M L 2004 *Random Matrices 3rd edn* (New York: Elsevier/Academic)
[28] Chen Y and Manning S M 1996 *J. Phys. A: Math. Gen.* **29** 7561
Chen Y and Manning S M 1994 *J. Phys. A: Math. Gen.* **27** 3615
[29] James A T 1964 *Ann. Math. Stat.* **35** 475
[30] Wigner E P 1951 *Proc. Camb. Phil. Soc.* **47, 790**
[31] Tracy C and Widom H 1994 *Commun. Math. Phys.* **159** 151
Tracy C and Widom H 1996 *Commun. Math. Phys.* **177** 727
For a review, see Tatsien L I 2002 *Proceedings of the International Congress of Mathematicians (Beijing 2002)* vol I (Beijing: Higher Education Press) pp 587–96
[32] Cavagna A, Garrahan J P and Giardina I 2000 *Phys. Rev.* B **61** 3960
[33] Fyodorov Y V 2004 *Phys. Rev. Lett.* **92** 240601
Fyodorov Y V 2005 *Acta Phys. Pol.* B **36** 2699
[34] Susskind L 2003 *Preprint* hep-th/0302219
Douglas M R, Shiffman B and Zelditch S 2004 *Commun. Math. Phys.* **252** 325
[35] Mersini-Houghton L 2005 *Class. Quantum Grav.* **22** 3481
[36] Aazami A and Easther R 2006 *J. Cosmol. Astropart. Phys.* JCAP03(2006)013
[37] Bray A J and Dean D S 2006 *Preprint* cond-mat/0611023
[38] Fyodorov Y V, Sommers H-J and Williams I *Preprint* cond-mat/0611585
[39] Kanzieper E 2002 *Phys. Rev. Lett.* **89** 250201
[40] Tricomi F G 1957 *Integral Equations (Pure Appl. Math V)* (London: Interscience)
[41] Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions* (New York: Dover)