Non-Hermitian random matrix theory: summation of planar diagrams, the 'single-ring' theorem and the disc–annulus phase transition

2006 J. Phys. A: Math. Gen. 39 10029

(http://iopscience.iop.org/0305-4470/39/32/S07)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 10029–10056 doi:10.1088/0305-4470/39/32/S07

Non-Hermitian random matrix theory: summation of planar diagrams, the 'single-ring' theorem and the disc–annulus phase transition

Joshua Feinberg

Physics Department, University of Haifa at Oranim, Tivon 36006, Israel
and
Physics Department, Technion, Israel Institute of Technology, Haifa 32000, Israel

E-mail: joshua@physics.technion.ac.il

Received 23 March 2006
Published 26 July 2006
Online at stacks.iop.org/JPhysA/39/10029

Abstract

I review aspects of work done in collaboration with A Zee and R Scalettar (1997 *Nucl. Phys.* B **504** 579; 1997 *Nucl. Phys.* B **501** 643; 2001 *J. Math. Phys.* **42** 5718) on complex non-Hermitian random matrices. I open by explaining why the bag of tools used regularly in analysing Hermitian random matrices cannot be applied directly to analyse non-Hermitian matrices, and then introduce the method of Hermitization, which solves this problem. Then, for rotationally invariant ensembles, I derive a master equation for the average density of eigenvalues in the complex plane, in the limit of infinitely large matrices. This is achieved by resumming all the planar diagrams which appear in the perturbative expansion of the Hermitized Green function. Remarkably, this resummation can be carried out *explicitly* for any rotationally invariant ensemble. I prove that in the limit of infinitely large matrices, the shape of the eigenvalue distribution is either a disc or an annulus. This is the celebrated 'single-ring' theorem. Which of these shapes is realized is determined by the parameters (coupling constants) which determine the ensemble. By varying these parameters a phase transition may occur between the two possible shapes. I briefly discuss the universal features of this transition. As the analysis of this problem relies heavily on summation of planar Feynman diagrams, I make a special effort at presenting a pedagogical exposition of the diagrammatic method, which some readers may find useful.

PACS numbers: 02.50.Cw, 11.10.Kk, 11.15.Pg, 71.20.−b, 71.23.An

1. Introduction

Non-Hermitian random matrices have interesting and important applications in various physical problems. Examples of such applications are the phase diagram of QCD [4, 5], nuclear decay and resonances in multichannel scattering in chaotic and disordered systems [6, 7], neural networks [7], and asymmetric random hopping matrices—the Hatano–Nelson model [8]—with applications to pinning of magnetic flux lines in high-temperature superconductors and also as a new tool for studying Anderson localization in disordered conductors. Studying Anderson localization by means of non-Hermitian matrices is also one of the reasons for the very recent growing interest in studying the statistics of resonance widths and delay times in random media [6, 10]. Our final example is diffusion in random velocity fields [11], which gives rise to random Fokker–Planck operators. For a nice recent overview of random matrix theory and its applications, including non-Hermitian matrices, see [12].

1.1. The difficulty posed by non-Hermitian matrix ensembles

Let us recall first a few basic facts about Hermitian random matrices. A basic tool in studying Hermitian random matrices ϕ (henceforth all matrices will be taken to be $N \times N$ with N tending to infinity unless otherwise specified) is the Green's function defined by

$$G(z) = \left\langle \frac{1}{N} \operatorname{Tr} \frac{1}{z - \phi} \right\rangle \tag{1}$$

where $\langle \cdots \rangle$ denotes averaging over the random distribution from which the matrices ϕ are drawn. Diagonalizing ϕ by a unitary transformation we have

$$G(z) = \left\langle \frac{1}{N} \sum_{k=1}^{N} \frac{1}{z - \lambda_k} \right\rangle$$

where the N real numbers $\{\lambda_k\}$ are the eigenvalues of ϕ. Thus, $G(z)$ is a meromorphic function with poles at the eigenvalues of ϕ. In the large-N limit, the poles merge into a cut (or several cuts) on the real axis[1]. Powerful theorems from the theory of complex-analytic functions can then be brought to bear to the problem of determining $G(z)$ [13]. All of this is well known.

In contrast, for a non-Hermitian matrix, the eigenvalues invade the complex plane. For example, for non-Hermitian matrices ϕ generated with the probability $P(\phi) = \frac{1}{Z} \mathrm{e}^{-N \operatorname{Tr} \phi^\dagger \phi}$, Ginibre determined long ago [14] that in the large-N limit, the eigenvalues are uniformly distributed over a disc of radius unity in the complex plane. The Green's function corresponding to this ensemble is

$$G(z, z^*) = \left\langle \frac{1}{N} \operatorname{Tr} \frac{1}{z - \phi} \right\rangle = \begin{cases} z^* & |z| < 1 \\ \dfrac{1}{z} & |z| > 1, \end{cases} \tag{2}$$

and it is obviously not analytic inside the unit disc. In general, the eigenvalues of a non-Hermitian matrix taken from some probability ensemble will occupy, on average, a two-dimensional domain $\mathcal{D}$ of the complex plane, where the Green's function is non-analytic. Thus, we lose the powerful aid of analytic function theory in analysing these ensembles.

In the theory of Hermitian random matrices, a powerful method consists of the Feynman diagrammatic expansion [15] in which one expands $G(z) = \sum_k \langle \operatorname{Tr} \phi^k \rangle / z^{k+1}$ as a series in

[1] More precisely, this is really a statement on the eigenvalues of an *individual* realization of ϕ in the large-N limit (assuming that all eigenvalues are typically confined in a finite segment, independent of N). The averaging process smudges the eigenvalues into cuts already at finite N.

$1/z$, the 'bare quark propagator'. This method is clearly no longer available in studying non-Hermitian random matrices. The knowledge of $G(z)$ as a series in $1/z$ can no longer tell us anything about the behaviour $G(z)$ for small z, as the simple example (2) above makes clear. The eigenvalues fill a two-dimensional region rather than a one-dimensional region, as is the case for Hermitian matrices.

A renormalization group inspired method used in [16] implicitly involves an expansion in $1/z$ and is thus also not available for dealing with non-Hermitian matrices without suitable further developments.

1.2. Some basic formalism and notation

I shall now derive a couple of standard formulae which play a central role in studying the eigenvalue distribution of random non-Hermitian matrices. The derivation is straightforward (see, e.g., [7] and section 2 in [1], and references therein), and is based on the identities[2] $\partial(1/z^*) = \partial^*(1/z) = \pi\delta(x)\delta(y)$ and $\partial\partial^* \log(zz^*) = \pi\delta(x)\delta(y)$.

We are now ready to deal with the average density of eigenvalues of a non-Hermitian matrix ϕ. We diagonalize the matrix by a similarity transformation[3]

$$\phi = S^{-1}\Lambda S \tag{3}$$

where Λ denotes a diagonal matrix with elements $\lambda_i, i = 1, \ldots, N$. Taking the Hermitian conjugate we have $\phi^\dagger = S^\dagger\Lambda^* S^{-1\dagger}$. By definition, the density of eigenvalues is

$$\rho(x, y) = \left\langle \frac{1}{N}\sum_i \delta(x - \operatorname{Re}\lambda_i)\delta(y - \operatorname{Im}\lambda_i) \right\rangle. \tag{4}$$

From the identities mentioned above, we obtain that $\partial^* \operatorname{Tr}(z - \Lambda)^{-1} \equiv \partial^* \operatorname{Tr}(z - \phi)^{-1} = \pi\sum_i \delta(x - \operatorname{Re}\lambda_i)\delta(y - \operatorname{Im}\lambda_i)$. Thus, using the identity $\det(z - \Lambda)(z^* - \Lambda^*) = \det(z - \phi)(z^* - \phi^\dagger)$ we may express ρ in terms of ϕ, and in a manner which is manifestly symmetric in z and z^*, as

$$\rho(x, y) = \frac{1}{\pi}\partial\partial^* \left\langle \frac{1}{N} \operatorname{Tr}\log(z - \phi)(z^* - \phi^\dagger) \right\rangle. \tag{5}$$

This expression involves the logarithm inside the average, as well as two derivatives. Thus, unless one has a simple way of calculating the average in (5) (e.g., the replica method in the case of Gaussian ensembles [7]), (5) is not a practical expression for $\rho(x, y)$. In many cases, it is easier to calculate Green's function

$$G(z, z^*) = \left\langle \frac{1}{N} \operatorname{Tr}\frac{1}{z - \phi} \right\rangle = \left\langle \frac{1}{N}\sum_i \frac{1}{(x - x_i) + \mathrm{i}(y - y_i)} \right\rangle, \tag{6}$$

(with $\lambda_i = x_i + \mathrm{i}y_i$) than to do the average $\left\langle \frac{1}{N} \operatorname{Tr}\log(z - \phi)(z^* - \phi^\dagger) \right\rangle$. Then, from (4) and from one of the identities mentioned above, we have

$$\rho(x, y) = \frac{1}{\pi}\partial^* G(z, z^*), \tag{7}$$

which is a simpler representation of $\rho(x, y)$ than (5).

These two representations for $\rho(x, y)$ and the relation between them can be interpreted by recognizing (5) as a two-dimensional Poisson equation for the electrostatic potential

[2] Our notation is standard: for $z = x + \mathrm{i}y$ we define $\partial \equiv \frac{\partial}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y})$ so that $\partial z = 1$. Clearly, $\partial z^* = 0$. Similarly, we define $\partial^* \equiv \frac{\partial}{\partial z^*} = \frac{1}{2}(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y})$, so that $\partial^* z^* = 1$.

[3] Non-diagonalizable matrices constitute a set of measure zero in the ensemble.

$-\frac{1}{2}\left\langle \frac{1}{N}\,\mathrm{Tr}\log(z-\phi)(z^*-\phi^\dagger)\right\rangle$ created by the charge density $\rho(x, y)$. This connection between eigenvalue distributions of complex matrices and two-dimensional electrostatics has been long known in the literature [7]. Continuing along this line, consider Green's function $G(z, z^*)$ in (6). If we define the electric field $\mathbf{E} = (\mathrm{Re}\, G, -\mathrm{Im}\, G)$ then from the definition of $G(z, z^*)$ we have

$$\mathbf{E}(\mathbf{x}) = \int \mathrm{d}^2 y\, \rho(\mathbf{y}) \frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2} \tag{8}$$

and thus (7) is simply the statement of Gauss' law for this electrostatic problem.

1.2.1. A quaternionic interlude and speculation. I end this section with a speculative remark, made for the first time in [1]. As was pointed out earlier, in the theory of random Hermitian matrices it is well known that it is much easier to work with $G(z) = \left\langle \frac{1}{N}\,\mathrm{Tr}\,\frac{1}{z-\phi}\right\rangle$ rather than with the density of eigenvalues $\rho(\mu)$ directly, since the power of analytic functions can be brought to bear on $G(z)$. Of course, one can go from $\rho(x)$ to $G(z)$ with the identity

$$\frac{1}{\pi}\,\mathrm{Im}\,\frac{1}{x-\mathrm{i}\epsilon} = \frac{1}{\pi}\frac{\epsilon}{x^2+\epsilon^2} \to \delta(x). \tag{9}$$

Here, for random non-Hermitian matrices our formula (5) gives the density of eigenvalues $\rho(x, y)$ directly, but this formula is awkward to work with because of the logarithm. Its direct analogue in the theory of random Hermitian matrices would be $\rho(x) = \frac{\partial}{\partial x}\left\langle \frac{1}{N}\,\mathrm{Tr}\,\theta(x-\phi)\right\rangle$, which would also be extremely awkward to work with. It would thus be desirable to define a quantity analogous to $G(z)$ and develop a method for calculating it. Reasoning along these lines we are led to an attempt to write a function of a quaternionic variable, in the same way that going from $\rho(x)$ to $G(z)$ we went from a function of a real variable to a function of a complex variable. Indeed, the obvious analogue of (9) is easy to find, namely,

$$\begin{aligned}\frac{1}{2j}\left[\frac{1}{(z-j\epsilon)|z-j\epsilon|} + \mathrm{i}\frac{1}{(z-j\epsilon)|z-j\epsilon|}\mathrm{i}\right] \\ = \frac{\epsilon}{|z-j\epsilon|^3} = \frac{\epsilon}{(x^2+y^2+\epsilon^2)^{3/2}} \to 2\pi\delta(x)\delta(y).\end{aligned} \tag{10}$$

Here $z = x + \mathrm{i}y$ is a complex number, ϵ is a small positive real number and $\{1, i, j, k\}$ is the standard basis of the quaternion algebra[4]. (The absolute value of a quaternion is defined by $|a + ib + jc + kd| \equiv \sqrt{a^2+b^2+c^2+d^2}$.)

Unfortunately, (10) as it stands, is less useful than (5), because it leads to the quaternionic Green's function $G(q)$ (q being a quaternionic variable)

$$G(q) = \left\langle \frac{1}{N}\sum_i \frac{1}{(q-\lambda_i)|q-\lambda_i|}\right\rangle \tag{11}$$

which involves the absolute value operation explicitly, and thus cannot be written as a simple trace like (1). Note, however, that if we split q quite generally into $q = z + jw$ (z, w being ordinary complex variables), then $G(z + jw)$ is manifestly non-analytic in z as $w \to 0$, even before taking the average.

This was posed in [1] as an interesting problem, namely to find a useful quaternionic generalization of (1) for random non-Hermitian matrices, and to develop the analogue of the work of Brézin *et al* [13] associated with it. Unfortunately, this problem is still open, at the moment of transcribing this talk.

[4] One may of course write an equivalent formula with j replaced everywhere by the third quaternion basis element k. Also, ϵ may be taken complex.

The rest of this paper is organized as follows. In section 2 I present the method of Hermitization [1, 5, 9, 11] by which the problem of determining the eigenvalue density of random non-Hermitian matrices can be reduced to the more familiar problem of determining the eigenvalue density of random Hermitian matrices. In section 3 I explain how this method was applied in [2] to derive a master formula for the density of eigenvalues of non-Hermitian random matrices taken from a large class of rotationally invariant non-Gaussian probability ensembles. This section contains a detailed pedagogical exposition of the diagrammatic method and its large-N behaviour, which is the basis for the discussion in that section. In section 4 I present numerical simulations of the quartic ensemble, carried in [3], and demonstrate that they fit well with the theoretical predictions of the formalism developed in section 3. In section 5 I formulate and prove the 'single-ring' theorem of [2], according to which, in the large-N limit, the shape of the eigenvalue distribution associated with any of the ensembles studied in section 3 is either a disc or an annulus. Finally, in sections 6 and 7, following [3], I present universal properties of the disc and annular phases and of the transition between them, as well as nice explicit expressions for the location of the boundaries of the eigenvalue distribution and its boundary values. Unfortunately, for lack of space, I cannot cover the interesting topic of adding random non-Hermitian matrices. The interested reader could read about it in [2].

2. The method of Hermitization: reduction to random Hermitian matrices

I have emphasized that a straightforward diagrammatic method is not allowed for non-Hermitian matrices. Somewhat remarkably, it is possible to arrive at a diagrammatic method indirectly. I will show that the problem of determining the eigenvalue density of random non-Hermitian matrices can be reduced to the problem of determining the eigenvalue density of random Hermitian matrices, for which the diagrammatic method may be applied. This is achieved by the *method of Hermitization*, as it was dubbed in [1]. Variants of this idea were presented independently in [5, 9, 11]. (Here we follow the convention of [1].)

We start with the representation $\rho(x, y) = \frac{1}{\pi}\partial\partial^*\left\langle\frac{1}{N}\operatorname{Tr}\log(z-\phi)(z^*-\phi^\dagger)\right\rangle$ (namely, equation (5)). By standard manipulations, we observe that $\langle\operatorname{Tr}_{(N)}\log(z-\phi)(z^*-\phi^\dagger)\rangle = \langle\operatorname{Tr}_{(2N)}\log H\rangle - i\pi N$, where H is the Hermitian $2N\times 2N$ matrix Hamiltonian

$$H = \begin{pmatrix} 0 & \phi - z \\ \phi^\dagger - z^* & 0 \end{pmatrix}. \tag{12}$$

Thus, if we can determine

$$F(\eta; z, z^*) = \frac{1}{2N}\langle\operatorname{Tr}_{(2N)}\log(\eta - H)\rangle \tag{13}$$

we can determine $\rho(x, y)$.

The matrix H is what is called a chiral matrix[5]. Due to its block structure, it anticommutes with the 'Γ_5' matrix $\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$, with the immediate consequence that its spectrum consists of pairs of eigenvalues with equal magnitudes and opposite signs. Namely, if ξ is an eigenvalue of H, so is $-\xi$.

Consider now the propagator matrix associated with H, namely,

$$\mathcal{G}_{\mu\nu}(\eta; z, z^*) = \left\langle\left(\frac{1}{\eta - H}\right)_{\mu\nu}\right\rangle \tag{14}$$

[5] The problem of determining Green's function of chiral matrices such as H has been discussed by numerous authors [17–19]. (See also the paper by Feinberg and Zee cited in [16].)

where η is a complex variable and the indices μ and ν run over all possible $2N$ values. Here we followed a common practice and borrowed some terminology from quantum chromodynamics: we may consider $\phi, \phi^\dagger$ as 'gluons' (in zero spacetime dimensions), which interact with a $2N$-dimensional multiplet of 'quarks' ψ^μ, with a complex mass matrix (the bare 'inverse propagator')

$$\mathcal{G}_0^{-1} = \begin{pmatrix} \eta & z \\ z^* & \eta \end{pmatrix} \tag{15}$$

(expressed in terms of its $N \times N$ blocks). $\mathcal{G}_{\mu\nu}$ is thus the propagator of these quarks in the fluctuating gluon field. The crucial point is that since H is Hermitian, $\mathcal{G}_{\mu\nu}$ can be determined by the usual methods of Hermitian random matrix theory. In particular, as we already mentioned, the diagrammatic evaluation of (14) is essentially the expansion of $\mathcal{G}_{\mu\nu}$ in powers of $1/\eta$, with interaction vertices H. This is a well-defined procedure for large η, and it converges to a unique function which is analytic in the complex η plane, except for the cut (or cuts) along the real axis which contain the eigenvalues of H. After summing this series (and thus determining $\mathcal{G}_{\mu\nu}(\eta; z, z^*)$ in closed form), we are allowed to set $\eta \to 0$ in (14). Speaking colloquially, we may say that the crucial manoeuvre here is that while we cannot expand in powers of z, we can arrange for z to 'hitch a ride' with η and at the end of the ride, throw η away.

We can now calculate $\rho(x, y)$ by two different methods, each with its advantages. The first is to observe that

$$-\frac{1}{H} = \begin{pmatrix} 0 & \frac{1}{z^* - \phi^\dagger} \\ \frac{1}{z - \phi} & 0 \end{pmatrix}.$$

Thus, the quantity $\frac{1}{z-\phi}$ is simply the lower left block of $\mathcal{G}_{\mu\nu}(\eta = 0; z, z^*)$. In other words, once we have $\mathcal{G}_{\mu\nu}(\eta; z, z^*)$ we can set η to zero and use (6) and (7) to write

$$\rho(x, y) = \frac{1}{N\pi} \partial^* \operatorname{Tr}_{(2N)} \left[\begin{pmatrix} 0 & \mathbf{1}_N \\ 0 & 0 \end{pmatrix} \mathcal{G}(0; z, z^*) \right]. \tag{16}$$

An alternative is to take the trace of $\mathcal{G}_{\mu\nu}$,

$$\mathcal{G}(\eta; z, z^*) = \frac{1}{2N} \left\langle \operatorname{Tr}_{(2N)} \frac{1}{\eta - H} \right\rangle = \frac{\eta}{N} \left\langle \operatorname{Tr}_{(N)} \frac{1}{\eta^2 - (z^* - \phi^\dagger)(z - \phi)} \right\rangle, \tag{17}$$

from which one can determine (13) (in terms of a dispersion integral), and therefore determine $\rho(x, y)$, as explained in [1] in detail. I shall not pursue this possibility here. However, a few important comments relating to (17) are in order. Observe from (17) that

$$\mathcal{G}(0 - \mathrm{i}0; z, z^*) = \frac{\mathrm{i}\pi}{N} \left\langle \operatorname{Tr}_{(N)} \delta\left(\sqrt{(z^* - \phi^\dagger)(z - \phi)}\right) \right\rangle \tag{18}$$

and thus counts the (average) number of zero-eigenvalues of the positive semi-definite Hermitian matrix $\sqrt{(z^* - \phi^\dagger)(z - \phi)}$. Thus, if (on average) ϕ has no eigenvalues equal to z, or in other words, if the density of eigenvalues of ϕ vanishes at z, then

$$\mathcal{G}(0 - \mathrm{i}0; z, z^*) = 0, \tag{19}$$

independently of the large-N limit. In particular, the boundaries of the eigenvalue distribution of ϕ in the large-N limit are the curve (or curves) in the complex plane which separate regions where $\mathcal{G}(0; z, z^*) = 0$ from regions where $\mathcal{G}(0; z, z^*) \neq 0$. As a matter of fact, one can infer the location of these boundaries even without an explicit knowledge of $\mathcal{G}(0; z, z^*)$ and $G(z, z^*)$, by investigating the 'gap equation' for $\mathcal{G}(0; z, z^*)$, namely, the trace of both sides of (14), and then setting $\eta = 0$ [1]. The 'gap equation' is a polynomial in $\mathcal{G}$ with real coefficients (which depend on z explicitly as well as through $G(z, z^*)$, and the parameters which appear

in the probability distribution for ϕ). Due to the chiral nature of H, this polynomial contains a trivial factor of $\mathcal{G}(0; z, z^*)$ which we immediately factor out. Setting $\mathcal{G} = 0$ in the remaining factor we obtain an equation for the boundary. This is discussed in detail in [1].

It is worth emphasizing that the formalism developed here has nothing to do with large N as such. It is also totally independent of the form of the probability distribution $P(\phi, \phi^\dagger)$. The formalism is of course indifferent to the method one may choose to use to determine $\mathcal{G}_{\mu\nu}(\eta; z, z^*)$.

To summarize, we have obtained a formalism—the 'method of Hermitization'—for reducing the problem of dealing with random non-Hermitian matrices to the well-studied problem of dealing with random Hermitian matrices. Given the non-Hermitian matrix ϕ, we study the Hermitian matrix H instead. By whatever method one prefers, once one has determined the quark propagator $\mathcal{G}_{\mu\nu}$ (or its trace, Green's function $\mathcal{G}(\eta; z, z^*)$), one can in principle obtain $\rho(x, y)$ using (16) (or using the other method mentioned following (17)). Whether that can be done in practice is of course another story. In the next section I shall present a large family of probability distributions $P(\phi, \phi^\dagger)$ for which it is possible to compute $\mathcal{G}_{\mu\nu}(\eta; z, z^*)$ in closed form (in the large-N limit) by a diagrammatic method. Therefore, for those models, the algorithm discussed above can in fact be followed all the way to the end.

3. Rotationally invariant non-Gaussian ensembles

The literature on random non-Hermitian prior to [2] has focused exclusively on Gaussian randomness, a prime example being Ginibre's work [14] on the probability distribution $P(\phi) = (1/Z)\exp(-N \operatorname{Tr} \phi^\dagger\phi)$. In [2] it was shown that using the method of Hermitization one can determine the density of eigenvalues of probability distribution of the form

$$P(\phi) = \frac{1}{Z}\, \mathrm{e}^{-N \operatorname{Tr} V(\phi^\dagger\phi)}, \tag{20}$$

where V is an arbitrary polynomial of its argument. Indeed, by a simple trick, it was shown in [2] that one can obtain the desired density of eigenvalues with a minimal amount of work, by judiciously exploiting the existing literature on random Hermitian matrices. In this section I explain how this can be done.

In a certain sense, the work presented in [2] may be thought of as the analogue of the work of Brézin *et al* for random Hermitian matrices [13]; they showed how the density of eigenvalues of Hermitian matrices ϕ taken from the probability distribution $P(\phi) = (1/Z)\exp[-N \operatorname{Tr} V(\phi)]$ with V an arbitrary polynomial can be determined, and not just for the Gaussian case studied by Wigner and others, in which $V = (1/2)\operatorname{Tr}\phi^2$. An important simplifying feature of the analysis in [13] is that $P(\phi)$ depends only on the eigenvalues of ϕ, and not on the unitary matrix that diagonalizes it. In contrast, the probability distribution (20) for non-Hermitian matrices depends explicitly on the $GL(N)$ matrix S used to diagonalize $\phi = S^{-1}\Lambda S$, and S does not decouple. Remarkably however, for the Gaussian $P(\phi)$, Ginibre [14] managed to integrate over S explicitly and derived an explicit expression for the probability distribution of the eigenvalues of ϕ. Unfortunately, it is not clear how to integrate over S and derive the expression for the eigenvalue probability distribution for non-Gaussian distributions of the form (20). In [2] this serious difficulty was circumvented by using the method of Hermitization.

Due to the symmetry of $P(\phi)$ under the transformation $\phi \to \mathrm{e}^{\mathrm{i}\alpha}\phi$, the density of eigenvalues is obviously rotational invariant. It is useful to introduce circularly invariant density as

$$\rho(x, y) \equiv \rho(r)/2\pi. \tag{21}$$

Rotational invariance thus leads to a simpler form of the defining formula (6) for $G(z, z^*)$, reproduced here as

$$G(z, z^*) = \left\langle \frac{1}{N} \operatorname{Tr} \frac{1}{z - \phi} \right\rangle = \int \mathrm{d}^2 x' \frac{\rho(x', y')}{z - z'},$$

which reads

$$\gamma(r) \equiv zG(z, z^*) = \int_0^r r' \, \mathrm{d}r' \rho(r'), \tag{22}$$

whence

$$\rho(r) = \frac{1}{r} \frac{\mathrm{d}\gamma}{\mathrm{d}r}. \tag{23}$$

Clearly, the quantity $\gamma(r)$ is a positive monotonically increasing function, which satisfies the obvious 'sum-rules'

$$\gamma(0) = 0 \qquad \text{and} \qquad \gamma(\infty) = 1. \tag{24}$$

In particular, observe that the first condition in (24) ensures that no $\delta(x)\delta(y)$ spike arises in $\rho(x, y)$ when calculating it from (7) with $G(z, z^*)$ given by (22), as it should be. We are now ready to perform the diagrammatic expansion for probability distributions of the form (20).

3.1. The diagrammatic expansion and its resummation

Consider the matrix Green's function $\mathcal{G}_{\mu\nu}$ in (14) and write it as

$$\mathcal{G}_{\mu\nu} = \left\langle \left(\frac{1}{\mathcal{G}_0^{-1} - V} \right)_{\mu\nu} \right\rangle, \tag{25}$$

where $\mathcal{G}_0^{-1}$ is the inverse propagator defined in (15) and V is the random chiral matrix (the 'gluon field')

$$V = \begin{pmatrix} 0 & \phi \\ \phi^\dagger & 0 \end{pmatrix}. \tag{26}$$

For large complex η we may perform the perturbative expansion of $\mathcal{G}_{\mu\nu}$ in powers[6] of $\mathcal{G}_0$,

$$\mathcal{G}_{\mu\nu} = \mathcal{G}_{0\mu\nu} + \langle (\mathcal{G}_0 V \mathcal{G}_0)_{\mu\nu} \rangle + \langle (\mathcal{G}_0 V \mathcal{G}_0 V \mathcal{G}_0)_{\mu\nu} \rangle + \cdots \tag{27}$$

which converges to a unique matrix function, analytic in the η-plane, cut along the appropriate segments on the real axis. This was our motivation to 'Hermitize' the problem to begin with.

The diagrammatic method [15] provides a convenient and efficient way of representing the various terms which appear in the perturbation series (27). Before sketching the Feynman rules for this expansion, some important symmetry considerations are in order.

As was mentioned following (14), $\mathcal{G}_{\mu\nu}$ is essentially the propagator for the quarks ψ^μ ($\mu = 1, 2, \ldots, 2N$) in the fluctuating gluon field V. Let us split the $2N$-dimensional quark multiplet ψ^μ into two N-dimensional 'flavours', $\psi = (u, d)$. We see then that the 'quark–quark–gluon' interaction is

$$\psi^\dagger V \psi = u^\dagger \phi d + \mathrm{d}^\dagger \phi^\dagger u. \tag{28}$$

This interaction term and the probability distribution (20) are manifestly invariant under the gauge transformation $\phi \to P\phi Q^\dagger$, $\phi^\dagger \to Q\phi^\dagger P^\dagger$, $u \to Pu$ and $d \to Qd$, where P and Q are

[6] Obviously, only terms with an even number of V survive in (27) for probability distributions of the form (20).

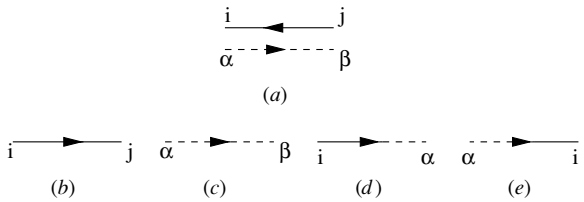

Figure 1. The bare gluon propagator (*a*), and the four blocks of the bare quark propagator (*b*)–(*e*). All propagators are of order $\frac{1}{N}$.

two independent unitary matrices, i.e., under the $U(N) \oplus U(N)$ subgroup of $U(2N)$. These transformations may be written more compactly as

$$V \to \mathcal{U} V \mathcal{U}^\dagger, \qquad \psi \to \mathcal{U}\psi, \qquad \text{where} \quad \mathcal{U} = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}. \tag{29}$$

We shall index quantities which transform under the first $U(N)$ in (29) by Latin indices $i, j, \ldots$, and quantities which transform under the second $U(N)$ by indices from the beginning of the Greek alphabet $\alpha, \beta, \ldots$.(Indices from the middle of the Greek alphabet, $\mu, \nu, \ldots$ are reserved for $2N$-dimensional objects.) Thus, for example, we have u_i, d_α, and the mixed object $\phi_{i\alpha}$.

3.1.1. Feynman rules and the topological nature of the $\frac{1}{N}$ expansion. Let us turn now to the Feynman rules for our diagrammatic expansion. Their presentation will be only sketchy and qualitative, since all that will really matter to us in the discussion below is counting powers of $1/N$. In fact, we are about to sum an infinite set of leading diagrams without ever computing any specific Feynman diagram.

We start with the propagators. The bare quark propagators $\langle u_i u_j^* \rangle$, $\langle u_i d_\alpha^* \rangle$, $\langle d_\alpha u_i^* \rangle$ and $\langle d_\alpha d_\beta^* \rangle$ are simply the four $N \times N$ blocks of $\frac{1}{N}\mathcal{G}_0$. They are all of order $\frac{1}{N}$. (Note that the object traced over in (16) and (17) is essentially $\frac{1}{N}\mathcal{G}_{\mu\nu}$. This trivial observation gives at least a practical justification for the factor $\frac{1}{N}$ in the bare quark propagator. A deeper reason for this normalization of the quark propagator will be provided below.) These propagators are indicated in figures 1(*b*)–(*e*). The diagonal blocks of $\frac{1}{N}\mathcal{G}_0$ correspond to quark propagation without flavour changing. The uu propagator is indicated in figure 1(*b*) by a full line, and the dd propagator is indicated in figure 1(*c*) by a dashed line. The off-diagonal blocks correspond to quark propagation with flavour changing. The lines are all directed since the quarks are 'charged' complex fields. Since all the blocks of $\frac{1}{N}\mathcal{G}_0$ are proportional to the $N \times N$ unit matrix, colour is preserved under quark propagation, whether flavour is changed or not.

Consider now the 'pure glue' action

$$N \operatorname{Tr} V(\phi^\dagger \phi) = N m^2 \operatorname{Tr} \phi^\dagger \phi + N g_4 \operatorname{Tr}(\phi^\dagger \phi)^2 + N g_6 \operatorname{Tr}(\phi^\dagger \phi)^3 + \cdots \tag{30}$$

in the exponent of (20). By definition, it is proportional to N, and the 'gluon self-couplings' $m^2, g_4, g_6, \ldots$ are all of order N^0. Thus, a typical matrix element in ϕ, drawn from the ensemble defined by (20), is of order $1/\sqrt{N}$ (and arbitrary phase), which leads, by standard arguments, to the conclusion that typical eigenvalues of ϕ are of order N^0, as was already mentioned in the context of Ginibre's ensemble.

The bare gluon propagator can be read off the quadratic term in (30) in the usual manner. We find

$$\langle \phi_{i\alpha} \phi_{j\beta}^* \rangle_0 = \frac{1}{N m^2} \delta_{ij} \delta_{\alpha\beta}. \tag{31}$$

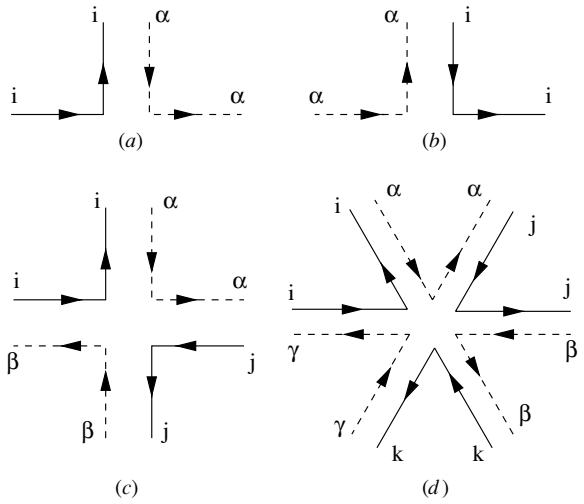

Figure 2. The two bare quark–gluon vertices (a) and (b) and quartic (c) and sextic (d) gluon self-interaction vertices. All vertices are of order N.

It is of order $\frac{1}{N}$, as a direct consequence of the explicit N-dependence in (20). In order that the quark propagator have the same N-dependence as the gluon's, we scaled the quark propagator to be of order $\frac{1}{N}$ as well. This behaviour, together with the fact that the 'gluon self-couplings' $m^2, g_4, g_6, \ldots$ are all of order N^0, will ensure that the model has a smooth behaviour as $N \to \infty$, as will be seen in more detail below.

The bare gluon propagator is depicted in figure 1(a). Following 't Hooft [20], it is represented by a double line, simply because ϕ carries two indices. The flows along these lines preserve both colour and flavour, as is evident from (31). The reason why the lines in figure 1(a) run in opposite directions is easy to understand. Observe that ϕ transforms under gauge transformations as $u \otimes d^\dagger$. Thus, the bare propagation of ϕ is similar to a process in one end of which a u-quark is destroyed and a d-quark is created, and in the other end of which the opposite occurs.

Let us discuss now the vertices. From the 'quark–quark–gluon' interaction (28) we see that a gluon ϕ always absorbs a d-quark and emits a u-quark, and $\phi^\dagger$ does the opposite. These two processes are described by the two quark–gluon vertices depicted in figures 2(a) and (b). Both vertices are taken to be proportional to N, so that they be of the same order of magnitude as the gluon self-interaction vertices, such as the quartic vertex with coupling Ng_4 shown in figure 2(c) and the sextic vertex with coupling Ng_6 in figure 2(d). Another way of seeing that the quark–gluon vertices should be of order N is to note from (25) that with the $\frac{1}{N}$ normalization of the quark propagator we have $\frac{1}{N}\mathcal{G}_{\mu\nu} = \left\langle \left(\frac{1}{N\mathcal{G}_0^{-1} - NV} \right)_{\mu\nu} \right\rangle$.

That all vertices scale in the same way with N, as do the propagators, is needed to obtain a smooth large-N behaviour. Note from figures 2(a)–(d) that both colour and flavour flow unobstructed through the various vertices. This is, of course, because all interaction terms are gauge invariant.

To summarize, all bare vertices are of order N, and all propagators are of order $\frac{1}{N}$.

Now, use the vertices and propagators defined above as building blocks of graphs—Feynman diagrams. The diagrams thus constructed are called double-line diagrams.

In order to compute a certain observable, all double-line diagrams consistent with that observable should be drawn, computed and added with the appropriate combinatoric factors.

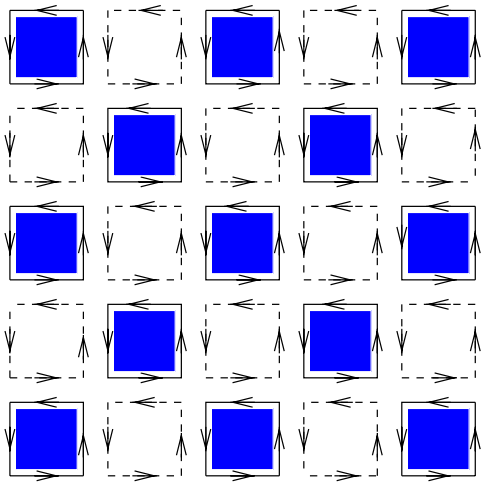

Figure 3. An example of a simply connected checkerboard surface which contains, in addition to quark–gluon vertices, only quartic gluon self-interaction vertices. There is a single quark loop which forms the edge. This diagram has 16 quark–gluon vertices plus 16 gluon quartic vertices ($V = 32$), 25 index loops (faces, $F = 25$), and 40 gluon propagators and 16 quark propagators (edges, $E = 56$). As explained below (see (32)), this diagram scales like N.

(This figure is in colour only in the electronic version)

Note that when a closed index loop occurs in a diagram, its index should be summed over. Therefore, we should assign a factor N to each index loop.

We would like now to study the $\frac{1}{N}$ behaviour of these double-line diagrams, and show that it is determined by their topological properties [20]. (An excellent discussion of this issue is given in [21], which I shall now follow with the appropriate modifications.) Let us start, for simplicity, with connected 'vacuum to vacuum' diagrams, i.e., connected double-line diagrams without external legs. (Such diagrams arise, for example, in the perturbative computation of the normalization factor Z in (20). Their sum is essentially $\log Z$. The expansion (27) of the quark propagator is of course not the sum of vacuum diagrams.) Since there are no external legs, each index line in the diagram has to be a closed *oriented* loop. Obviously, there are two types of such loops—loops made of a full line and loops made of a dashed line. Let us attach a little surface—a polygon—to each of these index loops. Since there are two kinds of loops, there are two kinds of such polygons. Thus, we can think of the double-line diagram in question as a tiled, triangulated (or more precisely polygonated) big two-dimensional surface, where any pair of edges of two different polygons, which share a gluonic double-line propagator, are identified. Since these two edges always run in opposite directions (recall figure 1(a)), the orientations of all polygons are the same, and the surface thus constructed is *orientable*[7]. Moreover, since the two lines which make a gluon propagator are of different types, and since index lines flow unobstructed through the vertices, any two neighbouring polygons which share a gluonic edge must be of different types. In other words, our surface has a checkerboard structure[8]. This surface may have edges, formed by quark lines. Also, it will generally be multiply connected, i.e., contain 'handles' and internal holes. An example of such a (simply connected) surface appears in figure 3.

[7] This is all due to the fact that all lines here are directed. For real *asymmetric* matrices, we would obtain unoriented lines and therefore unoriented surfaces [22], e.g., a Klein bottle.

[8] Such checkerboard surfaces appear in certain models of two-dimensional quantum gravity; see, e.g., [19].

Counting the power of N associated with such surface is easy. Assume the surface in question has V vertices, E edges and F faces. Each vertex is an interaction vertex in the Feynman diagram, and thus contributes a factor N. Each edge is a propagator which carries a factor $\frac{1}{N}$, and each face is an index loop carrying a factor N. All in all, the graph is proportional to $N^{F-E+V} \equiv N^{\chi}$, where

$$\chi = F - E + V \tag{32}$$

is the Euler character of the graph—a topological invariant of the graph.

As is well known, every connected oriented two-dimensional surface is topologically equivalent to a sphere with an appropriate number of punctures (holes) and handles (or 'wormholes') attached to it. For example, a torus is equivalent to a sphere with one handle attached, a disc is equivalent to a sphere with one puncture, and a cylinder is equivalent to a doubly punctured sphere, etc. Let us denote the number of handles (the so-called *genus* of the graph) by G, and the number of punctures (holes or boundaries) by b. Then a well-known theorem states that

$$\chi = 2 - 2G - b. \tag{33}$$

Thus, the leading vacuum diagrams in the limit $N \to \infty$ must have $G = b = 0$, i.e., the topology of a sphere. They are thus proportional to N^2. Obviously, they cannot contain quark loops, which are unpaired index loops and thus correspond to holes in the surface ($b \neq 0$). The last observation is more or less straightforward—one needs just one vertex factor to create either a pair of quarks or a triplet of gluons, but there are essentially N times more gluon types to sum over than quarks.

By removing a randomly chosen face off our leading-order graph, we can continuously stretch it open and project it onto the plane. For this reason, these leading vacuum graphs with spherical topology are also known as *planar* graphs. (We can also patch this planar graph back into a sphere, by including the point at infinity.)

Vacuum diagrams which depend of properties of quarks (i.e., on η, z, z^* in our case) must contain at least one quark loop. It follows from (33) that the leading such graphs in the large-N limit must contain precisely one quark loop (i.e., a single hole) and no handles. Such graphs have the topology of a disc, and could thus be drawn on the plane. They are planar graphs, like the pure glue graphs, with the sole difference between the two cases being that the outer boundary of that graph is that of the hole, i.e., the quark loop. There can be no quark loops in the *interior* of the graph in the large-N limit.

The reader may identify the Feynman rules described above as those which correspond to the zero-dimensional 'quantum field theory' with Lagrangian

$$\mathcal{L} = N \operatorname{Tr} V(\phi^\dagger \phi) + \psi^\dagger \left(N \mathcal{G}_0^{-1} - N V \right) \psi. \tag{34}$$

Diagrams which contribute in higher orders of the $\frac{1}{N}$ expansion of this theory do contain quark loops in their interiors, i.e., correspond to surfaces with internal holes. Diagrams with such internal quark loops obviously *do not* appear in the expansion of $\frac{1}{N}\langle \mathcal{G}_{\mu\nu} \rangle$ since the random matrix probability distribution (20) really does not contain any dynamical quarks. One method of suppressing these internal quark loops to any order in the $\frac{1}{N}$ expansion of (34) is to apply the so-called 'replica trick'[9]. Since here we are interested only in the $N \to \infty$ limit, namely, the leading order in the $\frac{1}{N}$ expansion, there is no need to invoke any replica considerations.

[9] That is to say, to replicate the quarks in (34) into n copies, with the same inverse bare propagator and coupling to the gluons. One then computes various observables in the replicated theory as a function of n, considered as a complex variable, and then take the limit $n \to 0$. Each internal quark loop costs a factor n and is therefore suppressed in that limit. (As an introductory reference on the replica trick see, e.g., A Zee's book cited in [15], chapter VI. 7.)

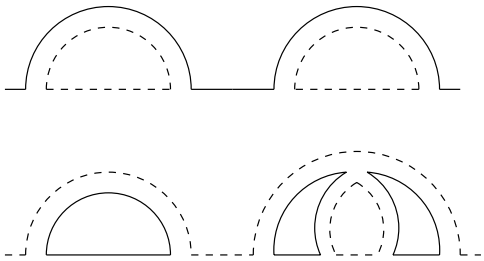

Figure 4. Two simple planar diagrams which contribute to $\frac{1}{N}\mathcal{G}_{\mu\nu}$. The upper one contributes to the uu block and the lower one to the dd block. Both diagrams are of order $\frac{1}{N}$. Colour indices and arrows are suppressed. These diagrams are one-quark reducible: their topology is such that they can be split into two disjoint pieces by cutting a single quark line.

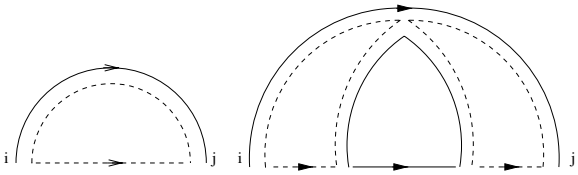

Figure 5. Two simple planar one-quark irreducible diagrams which contribute to the uu block of $N\Sigma_{\mu\nu}$: a diagram with no gluon self-interactions (a), and a diagram with a quartic gluon self-interaction (b). Evidently, both diagrams are proportional to δ_{ij}. Also, both diagrams are of order N.

3.1.2. Computation of $\frac{1}{N}\mathcal{G}_{\mu\nu}$ to leading order in $\frac{1}{N}$. Obviously, by cutting open the quark loop at the boundary of our vacuum diagram at one point, and attaching colour indices μ and ν at the two ends, we obtain a diagram which contributes to $\frac{1}{N}\mathcal{G}_{\mu\nu}$ in the expansion (27). Thus, to the leading order in $\frac{1}{N}$, all diagrams which contribute to $\frac{1}{N}\mathcal{G}_{\mu\nu}$ (and are therefore of order $\frac{1}{N}$) are planar diagrams, with the quark line running at the boundary, and all gluon double lines are attached to it on the *same side*, without crossing each other. A couple of such diagrams appear in figure 4.

I shall now explain how to sum all these planar diagrams systematically and thus obtain $\frac{1}{N}\mathcal{G}_{\mu\nu}$ in the limit $N \to \infty$.

The quark propagator (14) is given by the Dyson–Schwinger equation in terms of the one-quark irreducible self-energy $N\Sigma_{\mu\nu}$:

$$\frac{1}{N}\mathcal{G}_{\mu\nu} = \left(\frac{1}{N\mathcal{G}_0^{-1} - N\Sigma}\right)_{\mu\nu}. \tag{35}$$

We emphasize that this equation always holds, and practically amounts to the definition of the one-quark irreducible self-energy. Evidently, all planar self-energy diagrams are of order N.[10] Thus, to the leading order, $\Sigma_{\mu\nu}$ is of order N^0. A couple of planar one-quark irreducible diagrams which contribute to $N\Sigma_{\mu\nu}$ appear in figure 5.

[10] Their large-N behaviour is like that of the planar diagrams which contribute to $\frac{1}{N}\mathcal{G}_{\mu\nu}$ with their two external legs amputated. Amputating two external legs amounts to multiplication of the latter by N^2. Another way of looking at this is to say that the self-energy should scale with N as does the quark–quark–gluon vertex NV which is combined with the inverse propagator in the averaged quantity in (14).

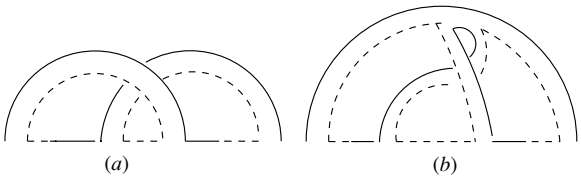

Figure 6. Two simple non-planar self-energy diagrams with a single crossing of gluon lines. Colour indices and arrows are suppressed. The effect of this crossing is dramatic: in diagram (*a*) there are no index loops at all (as opposed to its (reducible) counterpart in the upper part of figure 4 with two such loops), and in (*b*) there is only one index loop (as opposed to its planar counterpart in figure 5(*b*) with three such loops). These diagrams are both of order $\frac{1}{N}$, i.e., down by a factor N^{-2} with respect to the leading order.

For pedagogical clarity, we also display in figure 6 a couple of non-planar self-energy diagrams which contribute in the next-to-leading order.

The planar self-energy diagrams shown in figure 5 are clearly proportional to the unit matrix. By carefully studying a few additional planar self-energy diagrams the reader can convince himself or herself that this is a general property: in these planar self-energy diagrams, there is always a unique peripheral index line, which emanates from the quark–quark–gluon vertex at one end of the diagram, and participates in several gluon propagators and flows unobstructed through a number of gluon self-interaction vertices, without crossing any other index line, all the way to the quark–quark–gluon vertex at the other end of the diagram. (A slightly different way of looking at this is to connect the two ends of the diagram by a quark propagator and close it into a planar vacuum diagram. The point of gluing lies on a quark line, which is of course the *outer* edge of one of the polygons making the surface. The other edges of that polygon lie within the surface and combine to a directed index line, which becomes the peripheral index line mentioned above upon cutting the diagram open.) Thus, evidently, the off-diagonal blocks of $N\Sigma_{\mu\nu}$ are null and its uu and dd blocks must be proportional to the unit matrix. Furthermore, the u- and d-quark are completely equivalent. (We can switch the roles of ϕ and $\phi^\dagger$ in (20) and nothing would change.) Thus, the uu and dd blocks of $\Sigma_{\mu\nu}$ must be equal. In other words, for probability distributions of the form (20)

$$\Sigma_{\mu\nu} = \Sigma(\eta; r)\delta_{\mu\nu} \tag{36}$$

(where we recall that $r = |z|$.) Note that Σ depends on r and not separately on z and z^* because it may be expanded as an infinite sum of traces of $\mathcal{G}_0$, each term of which is a function only of η and r. Thus, from (35),

$$\mathcal{G}_{\mu\nu} = \frac{1}{r^2 - (\eta - \Sigma)^2} \begin{pmatrix} \Sigma - \eta & z \\ z^* & \Sigma - \eta \end{pmatrix}, \tag{37}$$

leading to

$$\mathcal{G} \equiv \frac{1}{2N} \sum_{\mu} \mathcal{G}_{\mu\mu} = \frac{\Sigma - \eta}{r^2 - (\eta - \Sigma)^2}. \tag{38}$$

The important point [2, 23] is that in the large-N limit, the one-quark irreducible self-energy $N\Sigma$ can be written in terms of the cumulants Γ_{2k} of $P(\phi) = (1/Z)\,\mathrm{e}^{-N\,\mathrm{Tr}\,V(\phi^\dagger\phi)}$ (equation (20)), namely, the connected correlators involving k ϕ's and $\phi^\dagger$'s. Thus, to the leading order in $\frac{1}{N}$, this cumulant is the sum of all connected planar pure glue diagrams with k external ϕ's and $\phi^\dagger$'s (with a prescribed index structure). By definition, Γ_{2k} is a quantity of order N. (It could be thought of essentially as the 'dressed' gluon self-interaction vertices Ng_{2k}.) This

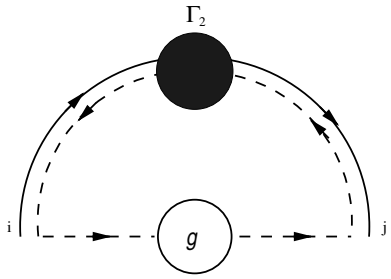

Figure 7. Contribution of the quadratic cumulant Γ_2 to self-energy.

makes sense, since any of the diagrams contributing to Γ_{2k} can be thought of as a planar, connected pure glue vacuum diagram (and thus of order N^2), cut open at k places along its *peripheral* gluon propagators[11] in such a way that it remains connected (i.e., there must be at least one gluon self-interaction vertex between consecutive cutting points), and then amputating the gluon legs thus formed. A moment's reflection reveals that this operation removes $k+1$ index loops (i.e., k internal index loops as well as the outer perimeter loop) and also k gluon propagators. Thus, it costs a factor of $\frac{1}{N}$, reducing the N-power counting of the diagram from 2 to 1.

Diagrammatically speaking, the Γ_{2k} may be thought of as a proper vertex of k ϕ's and $\phi^\dagger$'s, i.e., upon embedding it in a planar diagram it appears as a 'blob' out of which emanate k ϕ's and $\phi^\dagger$'s in an alternate manner and without crossings, and which cannot be separated into two smaller blobs, with k_1 ϕ's and $\phi^\dagger$'s in one blob and k_2 ϕ's and $\phi^\dagger$'s in the other (with $k_1 + k_2 = k$ of course).

The way $N\Sigma$ is determined by the cumulants Γ_{2k} can be deduced by working out a few simple cases. For example, the contribution of Γ_2 is shown in figure 7.

Counting powers of N in figure 7 is as follows: Γ_2 is of order N. The $\frac{1}{N}$ factors associated with the two gluon propagators poking out of the dark gluon blob representing Γ_2 are cancelled against the factors of N in the quark–gluon vertices. The horizontal line at the bottom is the *dressed* planar quark propagator, i.e., the sum of all planar diagrams contributing to $\frac{1}{N}\mathcal{G}_{\mu\nu}$, and is of order $\frac{1}{N}$. It is connected to the internal index line which emanates from the quark–quark–gluon vertex on the left and flows through the gluon dark blob. On the way, that line participates in several gluon propagators and flows unobstructed through a number of gluon self-interaction vertices in any of the planar pure glue diagrams which make the blob, without crossing any other index line, all the way to the quark–quark–gluon vertex on the right. Thus, it preserves its colour index, which must be summed over. Thus, the dressed quark propagator is traced over and produces a factor $\mathcal{G}(\eta; r)$ of order N^0, as defined in (38). The overall contribution of figure 7 to self-energy is therefore $\Gamma_2\mathcal{G}$ which is of order N. Evidently, the diagram shown in figure 5(*a*) is the leading contribution to this quantity. It will be instructive to analyse the contribution of Γ_4, the next cumulant, as well. Its contribution to self-energy is shown in figure 8.

Counting power of N in figure 8 is as follows: Γ_4 is of order N. The $\frac{1}{N}$ factors associated with the four gluon propagators poking out of the dark gluon blob are cancelled against the factors of N in the quark–gluon vertices. Each one of the three dressed quark propagators at the bottom is connected to an internal index loop which flows through the gluon dark blob.

[11] These cuts must be made along peripheral gluon propagators since the gluon diagram in question must be connected, without crossings, to other parts of a bigger planar diagram.

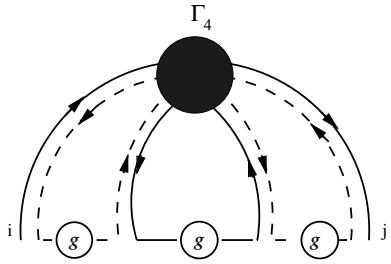

Figure 8. Contribution of the quartic cumulant Γ_4 to self-energy.

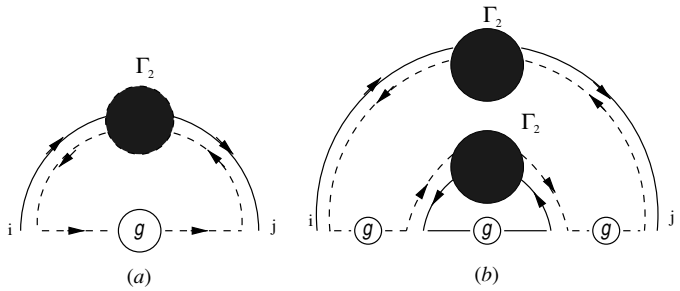

Figure 9. The contribution of the quadratic cumulant (*a*) and that of a separable gluon blob (*b*). The diagram (*b*) is evidently already accounted for in (*a*).

Thus it is traced over and produces a factor $\mathcal{G}(\eta; r)$. The overall contribution of figure 8 to self-energy is therefore $\Gamma_4\mathcal{G}^3$ which is of order N. Evidently, the diagram shown in figure 5(*b*) is the leading contribution to this quantity.

The general pattern should be obvious[12]. Thus, we obtain [2]

$$N\Sigma(\eta; r) = \sum_{k=1}^{\infty} \Gamma_{2k}[\mathcal{G}(\eta; r)]^{2k-1}, \tag{39}$$

where each of the summed terms is manifestly of order N.

The requirement, mentioned above, that the blob corresponding to Γ_{2k} be inseparable into smaller blobs (a property referred to as 'gluon connectedness' in [23]) can be explained with the help of the simple example shown in figure 9.

Evidently, the diagram in figure 9(*b*) is already accounted for by the contribution of the second cumulant. It is therefore not included in the contribution of the fourth cumulant shown in figure 8.

If we knew $\Sigma(\eta; r)$ in terms of $\mathcal{G}(\eta; r)$, then we can solve for $\mathcal{G}(\eta; r)$ and hence the density of eigenvalues. We are thus faced with the problem of determining the Γ_{2k} which appears to be a rather difficult task[13].

It is important to note that the Γ_{2k} in (39) depend only on the probability distribution $P(\phi)$ and not on the particular quantity we average over. Consider the problem of determining the eigenvalue density of the Hermitian matrix $\phi^\dagger\phi$. We would study the Green's function

$$F(w) = \left\langle \frac{1}{N} \operatorname{Tr}_{(N)} \frac{1}{w - \phi^\dagger\phi} \right\rangle \equiv \int_0^\infty \frac{\tilde{\rho}(\sigma)\, \mathrm{d}\sigma}{w - \sigma}, \tag{40}$$

[12] It should also be clear that we need not dress the quark–quark–gluon vertex, as such corrections are already accounted for by the terms summed in (39).

[13] Except, of course, for the Gaussian ensemble, for which only the variance $\Gamma_2 \neq 0$, and thus simply $\Sigma = \Gamma_2\mathcal{G}$.

where $\tilde{\rho}(\mu)$ is the averaged eigenvalue density of $\phi^\dagger\phi$. We see from (17) that this Green's function is related to $\mathcal{G}(\eta; r = 0)$ simply through

$$\mathcal{G}(\eta; r = 0) = \eta F(\eta^2). \tag{41}$$

Also, for $z = 0$ we have from (38)

$$\Sigma(\eta; 0) = \eta - \mathcal{G}^{-1}(\eta; 0). \tag{42}$$

The crucial observation is that $F(w)$ is already known in the literature on chiral and rectangular block random Hermitian matrices for the Gaussian distribution, as well as for non-Gaussian probability distributions of the form (20) with an arbitrary polynomial potential $V(\phi^\dagger\phi)$ [17–19]. (See also the paper by Feinberg and Zee cited in [16].) In fact, the authors of [18, 19] simply calculated the diagonal elements of the propagator $\mathcal{G}_{\mu\nu}(\eta; 0)$ using Dyson gas techniques (so the coefficients Γ_{2k} are only implicit in these papers).

The contents of (39), (41) and (42) may thus be summarized as follows: there is a unique function $a(\xi)$ which behaves like $1/\xi$ as ξ tends to infinity, is regular at $\xi = 0$, and satisfies the equation $\xi - a^{-1} = \sum_{k=1}^{\infty} \Gamma_{2k} a^{2k-1}$. This function is

$$a(\xi) = \xi F(\xi^2) \equiv \mathcal{G}(\xi; 0). \tag{43}$$

We, on the other hand, are interested in the opposite case, where $\eta = 0, r = |z| \neq 0$. Consider the matrix $\mathcal{G}_{\mu\nu}(0 - \mathrm{i}0; z, z^*)$. It follows from (35) and (36) that its $N \times N$ blocks are all proportional to the unit matrix $\mathbf{1}_N$ (we refer to this as 'colour index democracy'), and we see that

$$\mathcal{G}_{\mu\nu}(0 - \mathrm{i}0; z, z^*) = \begin{pmatrix} \mathcal{G} & G^* \\ G & \mathcal{G} \end{pmatrix}, \tag{44}$$

where $\mathcal{G}(0 - \mathrm{i}0; r)$ is given in (18) and the complex Green's function $G(z, z^*)$, the lower left block of $\mathcal{G}_{\mu\nu}(0 - \mathrm{i}0; z, z^*)$, was defined[14] in (6). Inverting the propagator in (44) and using (15) (with $\eta = 0$) and (35), we obtain the two equations

$$z = \frac{G^*}{|G|^2 - \mathcal{G}^2} \tag{45}$$

and

$$\Sigma = \frac{\mathcal{G}}{|G|^2 - \mathcal{G}^2}. \tag{46}$$

Note from (45) that if $\mathcal{G} = 0$, then $G = 1/z$, which corresponds to the region outside the domain of the eigenvalue distribution. In addition, using (45) twice, we obtain

$$zG = \frac{|G|^2}{|G|^2 - \mathcal{G}^2} = (|G|^2 - \mathcal{G}^2) r^2. \tag{47}$$

Thus, due to the fact that $\mathcal{G}$ is pure imaginary (see (18)) we conclude that $\gamma \equiv zG$ is always real and non-negative. Alternatively, we can write

$$\mathcal{G}^2 = \frac{zG(zG - 1)}{r^2}, \tag{48}$$

and therefore from the fact that $\mathcal{G}^2 \leqslant 0$ we conclude that $0 \leqslant \gamma \leqslant 1$, in accordance with the monotonicity of $\gamma(r)$ and the sum-rules $\gamma(0) = 0$ and $\gamma(\infty) = 1$ in (24).

Let us now define the pure imaginary quantity

$$\xi \equiv \frac{1}{\mathcal{G}} + \Sigma = \frac{1}{\mathcal{G}} + \sum_{k=1}^{\infty} \Gamma_{2k} \mathcal{G}^{2k-1}, \tag{49}$$

[14] From this point on, unless otherwise stated, our notation will be such that $\mathcal{G} \equiv \mathcal{G}(0 - \mathrm{i}0; r)$, $\Sigma \equiv \Sigma(0; r)$ and $G \equiv G(z, z^*)$. Note in particular from (18) that $\mathcal{G}$ is then pure imaginary, with $\mathrm{Im}\, \mathcal{G} \geqslant 0$.

where the last equality follows from (39) (with $\eta = 0$). Using (45) and (46) we thus have simply

$$\xi = \frac{zG}{\mathcal{G}}. \tag{50}$$

It then follows from (43) and the statement preceding that equation, and from (50) that

$$\xi F(\xi^2) = \mathcal{G} = \frac{zG}{\xi}. \tag{51}$$

We have thus eliminated the Γ_{2k}!

From (48) and from (50) we determine $\xi = \gamma/\mathcal{G}$ to be

$$\xi^2 = \frac{\gamma}{\gamma - 1} r^2. \tag{52}$$

Note that $\xi^2 \leqslant 0$, consistent with the bounds $0 \leqslant \gamma \leqslant 1$. Comparing (51) and (41) we obtain the remarkable relation

$$\mathcal{G}(\xi; 0) = \mathcal{G}(0; r), \tag{53}$$

for ξ and r that are related by (52) (with ξ imaginary).

We now have the desired equation for $\gamma \equiv zG$: substituting (52) into (51) we obtain

$$\gamma \left[r^2 F\left(\frac{\gamma r^2}{\gamma - 1} \right) - \gamma + 1 \right] = 0. \tag{54}$$

This is the master equation announced in the abstract. Thus, given F we can solve for $\gamma(r)$ using this equation.

From (40) we observe that for w large we may expand

$$F(w) = \frac{1}{w} + \frac{1}{w^2} f(w), \tag{55}$$

where $f(w) = f_0 + (f_1/w) + \cdots$. Substituting $w = \xi^2 = \gamma r^2/(\gamma - 1)$ into (55) and letting γ tend to 1, (54) becomes

$$(\gamma - 1)^2 \left[\frac{f(w)}{\gamma r^2} - 1 \right], \tag{56}$$

and thus $\gamma = 1$, which corresponds to the region outside the eigenvalue distribution, is always a double root of (54).

We will now explain how the Green's function $F(w)$ is obtained in the literature [18, 19]. (See also the paper by Feinberg and Zee cited in [16].) Let V be a polynomial of degree p. From the definition of $F(w)$ in (40) and its analyticity property, we expect F, following the arguments of Brézin *et al*, to have the form[15]

$$F(w) = \tfrac{1}{2} V'(w) - P(w)\sqrt{(w - a)(w - b)}, \tag{57}$$

where

$$P(w) = \sum_{k=-1}^{p-2} c_k w^k. \tag{58}$$

The constants $0 \leqslant a < b$ and c_k are then determined completely by the requirement that $F(w) \to \frac{1}{w}$ as w tends to infinity, and by the condition that $F(w)$ has at most an integrable singularity as $w \to 0$. (Thus, if $a > 0$, inevitably $c_{-1} = 0$. However, if $a = 0$, then c_{-1} will be determined by the first condition.)

[15] Here we assume for simplicity that the eigenvalues of $\phi^\dagger \phi$ condense into a single segment $[a, b]$.

4. An example: the quartic ensemble

Having determined $F(w)$ in this way, we substitute it into (54) and find $G(z, z^*)$. Equation (54) is an algebraic equation for $\gamma(r)$ and thus may have several r-dependent solutions. In constructing the actual $\gamma(r)$ one may have to match these solutions smoothly into a single function which increases monotonically from $\gamma(0) = 0$ to $\gamma(\infty) = 1$ (recall (24)). (An explicit non-trivial example of such a sewing procedure is the construction of $\gamma(r)$ in the disc phase of the quartic ensemble [2].) We can thus calculate the density of eigenvalues explicitly for an arbitrary V.

The Gaussian case with $V = \phi^\dagger\phi$ provides the simplest example. Here we have $2\xi F(\xi^2) = \xi - \sqrt{\xi^2 - 4}$, whence the roots of (54) are $\gamma = 0, 1$ and r^2. We note that $\gamma = 0$ is unphysical, $\gamma = r^2$ (i.e., $G = z^*$) corresponds to Ginibre's disc [14], and $\gamma = 1$ is the solution outside the disc.

As an explicit non-trivial example, the quartic ensemble with $V(\phi^\dagger\phi) = 2m^2\phi^\dagger\phi + g(\phi^\dagger\phi)^2$ was studied in detail in [2]. As should perhaps be expected in advance, the following behaviour in the parameter space $m^2, g > 0$ was found: for m^2 positive, the eigenvalue distribution was disc-like (and non-uniform), generalizing Ginibre's work, but as $m^2 \equiv -\mu^2$ was made more and more negative, a phase transition at the critical value $\mu_c^2 = \sqrt{2g}$ occurred, after which the disc fragmented into an annulus. The density of eigenvalues was calculated in [2] in detail. In particular, the radii of the disc and annulus were calculated explicitly in terms of the couplings in V. An interesting behaviour of the density of eigenvalues was observed through the disc–annulus phase transition. As the transition was approached from the disc side (e.g., by lowering m^2 towards $-\mu_c^2$ holding g fixed), a central region of *finite* radius in the disc got progressively depleted. At the transition depletion was complete and an annulus of finite critical radius was formed, with vanishing eigenvalue density at the inner radius of the annulus. As the couplings were tuned deeper into the annular phase, a finite jump in the density of eigenvalues at the inner radius appeared. The interested reader could find all relevant formulae in section 5 of [2] and section 4 of [3]. In what follows we shall compare these analytical large-N results, following [3], against numerical simulations.

In figure 10 we display the numerical results for $\rho_{\text{disc}}(r)$ for (128×128)-dimensional matrices, and compare them to the analytical large-N result (5.8) of [2] (or (4.4) of [3]). (As a trivial check of our numerical code, we also included in this figure the results for the Gaussian (Ginibre) ensemble.) Evidently, the agreement between the numerical and the analytical results is good. Note the finite-N effects near the edge of the disc. In figures 11(*a*)–(*c*) we display our numerical results for $\rho_{\text{annulus}}(r)$ for matrices of various sizes, and compare them to the analytical large-N result (5.18) of [2] (or (4.13) of [3]). In these figures we hold $-m^2 = \mu^2 = 0.5$ fixed, and increase g from 0.025 to 0.1. (Here we have $\mu^2 = 0.5 = \mu_c^2/2\sqrt{2g}$. Thus increasing g as indicated in the text brings us closer to the disc–annulus phase transition.)

In [3] we have also studied numerically the disc–annulus phase transition. We have measured the density of eigenvalues $\rho(r)$ of matrices ϕ of size 128×128, taken from the the quartic ensemble with $\mu^2 = 0.5$ and for $g = 0.025, 0.05, 0.1, 0.125, 0.15$ and 0.175. The results are displayed in figure 12.

For these values of g, we start in the annular phase at the lowest value of g. For our set of parameters we have $\mu^2 = 0.5 = \mu_c^2/2\sqrt{2g}$. Thus, increasing g (while keeping μ^2 fixed at 0.5) brings us closer to the disc–annulus phase transition, which occurs (at large N) at $g_c = 0.125$. Increasing g beyond that puts us into the disc phase.

The first three profiles on the right in figure 12 belong to the annular phase. As g increases towards the transition point at $g_c = 0.125$, these three graphs exhibit decrease of the inner radius of the annulus, in accordance with the analytical results.

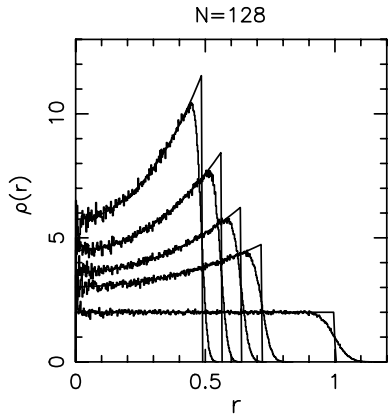

Figure 10. Comparison between Monte Carlo measurements of the density of eigenvalues $\rho(r)$ of matrices ϕ of size 128×128, taken from the quartic ensemble $V(\phi^\dagger\phi) = 2m^2\phi^\dagger\phi + g(\phi^\dagger\phi)^2$ with $m^2 = 0.5$ (disc phase) and for $g = 0, 0.5, 1, 2, 4$ (g increases from bottom to top), compared to the analytical results of [2] (solid lines). At $g = 0$ we obtain Ginibre's Gaussian ensemble with $V = \phi^\dagger\phi$, with its unit disc of eigenvalues. (This figure is taken from [3].)

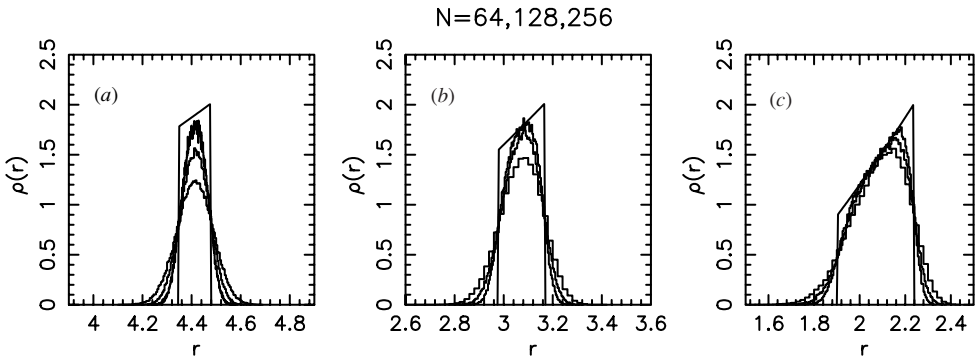

Figure 11. Results of Monte Carlo measurements of the density of eigenvalues $\rho(r)$ of matrices ϕ of sizes corresponding to $N = 64, 128$ and 256, taken from the quartic ensemble with $m^2 = -\mu^2 = -0.5$ (annular phase) for various values of the quartic coupling: $g = 0.025$ in (a), $g = 0.05$ in (b) and $g = 0.1$ in (c). These are compared to the analytical results of [2] (solid lines). As N increases, the numerical results converge monotonically to the analytical results. (This figure is taken from [3].)

The critical density profile, corresponding to $g_c = 0.125$, is the fourth profile (from the right). For our choice of parameters, the theoretical boundary radii of the critical annulus, i.e., at $g = 0.125$, are $R_{\text{in}}^{\text{crit}} = 1/\mu_c = \sqrt{2}$ and $R_{\text{out}}^{\text{crit}} = \sqrt{2}/\mu_c = 2$. These boundary values fit nicely with the features of the critical profile in figure 12.

Finally, the last two profiles in figure 12 have pronounced tails extending to $r = 0$ and thus belong to the disc phase.

5. Boundaries and shape universality of the eigenvalue distribution: the single-ring theorem

A remarkable property of (54) is that it has only two r-independent solutions: $\gamma = 0$ and $\gamma = 1$ [2]. Since the actual $\gamma(r)$ increases monotonically from $\gamma(0) = 0$ to $\gamma(\infty) = 1$,

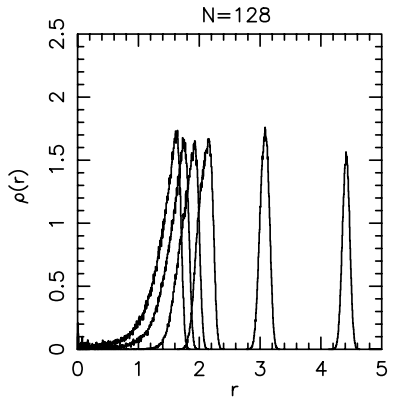

Figure 12. Monte Carlo measurements of the density of eigenvalues $\rho(r)$ of matrices ϕ of size 128×128, taken from the the quartic ensemble with $\mu^2 = 0.5$ and for $g = 0.025, 0.05, 0.1, 0.125, 0.15$ and 0.175 (g increases from right to left). The first three profiles on the right (corresponding to the three lowest values of g) evidently belong to the annular phase. The fourth density profile from the right is the critical one (corresponding to $g_c = 0.125$). Finally, the last two profiles (which correspond to the two higher values of g) belong to the disc configuration. (This figure is taken from [3].)

without any plateaux in between, we immediately conclude from this observation that there can be no more than a single void in the eigenvalue distribution. Thus, in the class of models governed by $P(\phi) = \frac{1}{Z}\,\mathrm{e}^{-N\,\mathrm{Tr}\,V(\phi^\dagger\phi)}$ (equation (20)), the shape of the eigenvalue distribution is either a disc or an annulus, whatever polynomial the potential $V(\phi^\dagger\phi)$ is. This result is the 'single-ring theorem' of [2].

The 'single-ring theorem' may appear counterintuitive at first sight. Indeed, consider a potential $V(\phi^\dagger\phi)$ with several wells or minima. For deep enough wells, we expect the eigenvalues of $\phi^\dagger\phi$ to 'fall into the wells'. Thus, one might suppose the eigenvalue distribution of ϕ to be bounded by a set of concentric circles of radii $0 \leqslant r_1 < r_2 < \cdots < r_{n_{\max}}$, separating annular regions on which $\rho(r) > 0$ form voids (annuli in which $\rho(r) = 0$). *A priori*, it is natural to assume that the maximal number of such circular boundaries should grow with the degree of V, because V may then have many deep minima. Remarkably, however, according to the 'single-ring theorem' the number of these boundaries is 2 at the most.

To reconcile this conclusion with the *a priori* expectation just mentioned, note that while the eigenvalues of the Hermitian matrix $\phi^\dagger\phi$ may split into several disjoint segments along the positive real axis, this does not necessarily constrain the eigenvalues of ϕ itself to condense into annuli. Indeed, the Hermitian matrix $\phi^\dagger\phi$ can always be diagonalized $\phi^\dagger\phi = U^\dagger\Lambda^2 U$ by a unitary matrix U, with $\Lambda^2 = \mathrm{diag}\big(\lambda_1^2, \lambda_2^2, \ldots, \lambda_N^2\big)$, where the λ_i are all real. This implies that $\phi = V^\dagger\Lambda U$, with V a unitary matrix as well. Thus, the complex eigenvalues of ϕ are given by the roots of $\det(z - \Lambda W) = 0$, with $W = UV^\dagger$. Evidently, as W ranges over $U(N)$ (which is what we expect to happen in the generic case), the eigenvalues of ΛW could be smeared (in the sense that they would not span narrow annuli around the circles $|z| = |\lambda_i|$.)

The last argument in favour of the 'single-ring theorem' clearly breaks down when W fails to range over $U(N)$, which occurs when the unitary matrices U and V are correlated. For example, ϕ may be such that $W = UV^\dagger$ is block diagonal, with the upper diagonal block being a $K \times K$ unitary diagonal matrix $\mathrm{diag}(\mathrm{e}^{\mathrm{i}\omega_1}, \ldots, \mathrm{e}^{\mathrm{i}\omega_K})$ (and with K a finite fraction of N). In the extreme case $K = N$, in which W is completely diagonal, $W \equiv \mathrm{e}^{\mathrm{i}\omega} = \mathrm{diag}(\mathrm{e}^{\mathrm{i}\omega_1}, \ldots, \mathrm{e}^{\mathrm{i}\omega_N})$, we see that $\phi = U^\dagger\,\mathrm{e}^{\mathrm{i}\omega}\Lambda U$ is a *normal* matrix (that is, $[\phi, \phi^\dagger] = 0$) with eigenvalues

$\mathrm{diag}(e^{i\omega_1}\lambda_1, \ldots, e^{i\omega_N}\lambda_N)$. Thus, normal matrices, or partially normal matrices (i.e., the case $K < N$), evade the 'single-ring' theorem: if the first K eigenvalues $\lambda_1^2, \lambda_2^2, \ldots, \lambda_K^2$ of $\phi^\dagger\phi$ split into several disjoint segments along the positive real axis, the corresponding eigenvalues of ϕ will split into concentric annuli in the complex plane obtained by revolving those λ-segments. Normal, or partially normal matrices are, of course, extremely rare in the ensembles of non-Hermitian matrices studied in this paper, and do not affect the 'single-ring' behaviour of the bulk of matrices in the ensemble.

Let us sketch now the proof of the single-ring theorem, following [2]. Thus, let us assume for the moment that the domain of eigenvalues has n boundaries. It is easy to see from (22), the defining equation of $\gamma(r)$, that in the annular void $r_k < r < r_{k+1}$ in the eigenvalue distribution, γ is a constant which is equal to the fraction of eigenvalues contained inside the disc $r \leqslant r_k$. Thus, the equation to determine γ (equation (54) which we repeat here for convenience),

$$\gamma\left[r^2 F\left(\frac{\gamma r^2}{\gamma - 1}\right) - \gamma + 1\right] = 0, \tag{59}$$

must have a series of monotonically increasing constant solutions $\gamma_1 < \gamma_2 < \cdots \leqslant 1$, which correspond to the various voids.

In particular, from (24) we have $\gamma = 1$ for $r > r_{n_{\max}}$, namely, $G = 1/z$. Thus, $\gamma = 1$ must be a solution of (59), the maximal allowed constant solution, as we already saw following (56). Also, if $r_1 > 0$, namely, if there is a hole at the centre of the eigenvalue distribution, then for $r \leqslant r_1$ we must have $\gamma = 0$ (independently of r), which is obviously a solution of (59). (From the paragraph right below (57) it is clear that $wF(w) \to 0$ at $w = 0$. Thus $\gamma = 0$ is indeed a solution. On the other hand, if $r_1 = 0$, so that the eigenvalue distribution includes the origin, an r-independent solution $\gamma = 0$ is of course only a spurious solution which should be discarded.)

Assume now that $\gamma = \gamma_0$ is an r-independent solution of (59). Taking the derivative of (59) with respect to r^2 at $\gamma \equiv \gamma_0$ we obtain

$$\gamma_0\left[F(\xi^2) + \xi^2\frac{\mathrm{d}F}{\mathrm{d}\xi^2}\right]\Bigg|_{\xi^2 = \frac{\gamma_0}{\gamma_0 - 1}r^2} = 0, \tag{60}$$

which is the condition for the existence of an r-independent solution γ_0. Thus, there are two possibilities: either $\gamma_0 = 0$, which we already encountered, or $F(\xi^2) + \xi^2\frac{\mathrm{d}F}{\mathrm{d}\xi^2} = 0$. This equation immediately yields

$$F(\xi^2) = \frac{1}{\xi^2}, \tag{61}$$

where the integration constant is fixed by the asymptotic behaviour of $F(\xi^2)$ as $\xi \to \infty$. But for a generic $V(\phi^\dagger\phi)$, $F(w)$ is given by (equation (40))

$$F(w) = \left\langle \frac{1}{N}\,\mathrm{Tr}_{(N)}\frac{1}{w - \phi^\dagger\phi}\right\rangle \equiv \int_0^\infty \frac{\tilde{\rho}(\sigma)\,\mathrm{d}\sigma}{w - \sigma},$$

with (61) being the asymptotic behaviour of $F(w)$ as $w = \xi^2 \to \infty$. We thus conclude that $\xi^2 \to \infty$ in (61), namely, that $\gamma_0 = 1$. Thus, to summarize, the only possible r-independent solutions of (54) are $\gamma = 0$ and $\gamma = 1$, which we already discussed. Since no other r-independent solutions arise, there can be no more than a single void in the eigenvalue distribution, whatever polynomial the potential $V(\phi^\dagger\phi)$ is. The shape of the eigenvalue distribution is thus either a disc or an annulus.

Interestingly enough, we can arrive at the same conclusion by invoking other general aspects of the method of Hermitization, and thus providing a nice self-consistency check of

our formalism. Recall at this point that the boundaries of the eigenvalue distribution are given by (19), namely, the zeros of $\mathcal{G}$. But $\mathcal{G}$ is given by (equation (48))

$$\mathcal{G}^2 = \frac{\gamma(\gamma-1)}{r^2},$$

from which we see that at the boundaries γ can take on only two values: 0 and 1. Since γ is a constant in the void, by continuity, these are the only possible values of γ inside any of the voids. Therefore, there may be two circular boundaries at most. In other words, as far as the eigenvalue density of ϕ is concerned, an ensemble of the form (20) may have two phases at most, as we concluded earlier.

In addition to $\gamma = 1$ (and possibly $\gamma = 0$), (59) must have other roots which do depend on r. Among all these other roots, we expect to find a unique root $\gamma(r)$, which is a positive monotonically increasing function of r, that matches continuously at the boundaries r_k ($k \leqslant 2$) to the constant roots of (59). The actual $zG(z, z^*)$ of the ensemble (20) is therefore a continuous monotonically increasing function of r, as required by (22). If the eigenvalue distribution is annular with boundaries $0 \leqslant r_1 < r_2$ (the case $r_1 = 0$ corresponding to the full disc), it vanishes for $0 \leqslant r \leqslant r_1$, rises monotonically from 0 to 1 on $r_1 \leqslant r \leqslant r_2$, and remains equal to 1 for $r \geqslant r_2$.

The sextic potential $V(\phi^\dagger\phi) = m^2\phi^\dagger\phi + \frac{\lambda}{2}(\phi^\dagger\phi)^2 + \frac{g}{3}(\phi^\dagger\phi)^3$ is the potential of the lowest degree in (20) for which the eigenvalues of $\phi^\dagger\phi$ may split into more than a single segment. Therefore, it is the simplest case in which the single-ring theorem has non-trivial consequences. In fact, it is easy to see that there can be at most two eigenvalue segments in the spectrum of $\phi^\dagger\phi$, in which case one segment touches the origin. In [3] we used Monte Carlo simulations to generate an ensemble of random matrices (of size 32×32) corresponding to (20) with the sextic potential. The coefficients of the latter were tuned to the phase where the spectrum of $\phi^\dagger\phi$ splits into two segments. We verified numerically that the eigenvalue distribution of ϕ in this phase was always a disc, and not a disc surrounded by a concentric annulus.

6. Universal features of the disc and annular phases and the transition between them

According to the 'single-ring' theorem, the eigenvalue distribution of ϕ is either a disc or an annulus. The behaviour of $F(w)$ as $w \sim 0$ turns out to be an indicator as to which phase of the two the system is in [3], as I now show.

(A) *Disc phase*. In the disc phase we expect that $\rho(0) > 0$, as in Ginibre's case. Thus, from (23) $\rho(r) = (1/r)(\mathrm{d}\gamma/\mathrm{d}r) \equiv 2(\mathrm{d}\gamma/\mathrm{d}r^2)$ and from the first sum-rule $\gamma(0) = 0$ in (24) we conclude that

$$\gamma(r) \sim \tfrac{1}{2}\rho(0)r^2 \tag{62}$$

near $r = 0$. Therefore, for r small, (54) yields

$$F\left(-\frac{\rho(0)r^4}{2} + \cdots\right) \sim -\frac{1}{r^2}, \tag{63}$$

namely, $F(w) \sim 1/\sqrt{w}$ for $w \sim 0$, as we could have anticipated from Ginibre's case. This means that in the disc phase we must set $a = 0$ in (57). Consequently, in the disc phase c_{-1} does not vanish. We can do even better: paying attention to the coefficients in (57) and (58) (with $a = 0$) we immediately obtain from (63) that

$$c_{-1} = \sqrt{\frac{\rho(0)}{2b}}. \tag{64}$$

(B) *Annular phase*. In the annular phase $\gamma(r)$ must clearly vanish identically in the inner void of the annulus. Thus, (54) implies that $F(w)$ cannot have a pole at $w = 0$, and therefore

from (57) we must have $c_{-1}\sqrt{ab} = 0$. Thus, the annulus must arise for $c_{-1} = 0$ (the other possible solution $a = 0, c_{-1} \neq 0$ leads to a disc configuration with $\gamma = 0$ only at $r = 0$, as we just discussed.)

Thus, to summarize, in the disc phase F has the form

$$F_{\rm disc}(w) = \frac{1}{2}V'(w) - \left(\sqrt{\frac{\rho(0)}{2b}}w^{-1} + c_0 + c_1 w + \cdots + c_{p-2}w^{p-2}\right)\sqrt{w(w-b)}, \tag{65}$$

while in the annular phase it has the form

$$F_{\rm annulus}(w) = \tfrac{1}{2}V'(w) - (c_0 + c_1 w + \cdots + c_{p-2}w^{p-2})\sqrt{(w-a)(w-b)}. \tag{66}$$

Having determined $F(w)$ in this way, i.e., having determined the various unknown parameters in (65) or in (66), we substitute it into (54) and find $G(z, z^*)$. We can thus calculate the density of eigenvalues $\rho(r)$ explicitly for an arbitrary V.

(C) *The disc–annulus phase transition*. We now turn to the disc–annulus phase transition. An important feature of this transition is that $F(w)$ is continuous through it. To see this, we argue as follows. By tuning the couplings in V, we can induce a phase transition from the disc phase into the annular phase, or vice versa. Note, of course, that we can parametrize any point in the disc phase either by the set of couplings in V or by the set of parameters $\{c_{-1}, c_0, \ldots, c_{p-2}; b\}$ in (65). The 'coordinate transformation' between these two sets of parameters is encoded in the asymptotic behaviour of $F(w)$. Similarly, we can parametrize any point in the annular phase either by the set of couplings in V or by the set of parameters $\{c_0, c_1, \ldots, c_{p-2}; a, b\}$ in (66). Due to the one-to-one relation (in a given phase, once we have established it is the stable one) between the couplings in V and the parameters in $F(w) - \frac{1}{2}V'(w)$ (namely, the c_n and the locations of the branch points of $F(w)$), we can describe the disc–annulus transition in terms of the latter parameters (instead of the couplings in V). Clearly, the transition point is reached from the disc phase when $\rho(0) = 0$, that is, when c_{-1} in (65) vanishes:

$$c_{-1}^{\rm crit} = 0. \tag{67}$$

Similarly, the transition point is reached from the annular phase when the lower branch point a in (66) vanishes. Thus, for example, in a transition from the disc phase into the annular phase, $F_{\rm disc}(w)$ in (65) would crossover continuously into $F_{\rm annulus}(w)$ in (66) through a critical form

$$F_{\rm crit}(w) = \tfrac{1}{2}V'_{\rm crit}(w) - \left(c_0^{\rm crit} + c_1^{\rm crit}w + \cdots + c_{p-2}^{\rm crit}w^{p-2}\right)\sqrt{w(w-b^{\rm crit})}. \tag{68}$$

The continuity of $F(w)$ through the transition was demonstrated explicitly in [2] for the quartic ensemble $V(\phi^\dagger\phi) = 2m^2\phi^\dagger\phi + g(\phi^\dagger\phi)^2$.

This discussion obviously generalizes to cases when $F(w)$ has multiple cuts, which correspond to condensation of the eigenvalues of $\phi^\dagger\phi$ into many segments. If $w = 0$ is a branch point of $F(w)$, that is, if the lowest cut extends to the origin, we are in the disc phase,

$$F_{\rm disc}(w) = \tfrac{1}{2}V'(w) - (c_{-1}w^{-1} + c_0 + c_1 w + \cdots + c_{p-2}w^{p-2})\sqrt{w(w-b_1)\cdots(w-b_n)}, \tag{69}$$

with $0 < b_1 < \cdots < b_n$. Relation (64) then generalizes to

$$c_{-1} = \sqrt{\frac{\rho(0)}{2(-1)^{n+1}\prod_{k=1}^{n} b_k}}. \tag{70}$$

Since c_{-1} must be real we conclude that such a configuration exists only for n odd.

If the lowest branch point in $F(w)$ is positive, we are in the annular phase with

$$F_{\text{annulus}}(w) = \tfrac{1}{2}V'(w) - (c_0 + c_1 w + \cdots + c_{p-2}w^{p-2})\sqrt{(w-a)(w-b_1)\cdots(w-b_n)}. \quad (71)$$

The phase transition would occur when the couplings in $V(\phi^\dagger\phi)$ are tuned such that $F_{\text{disc}}(w)$ and $F_{\text{annulus}}(w)$ match continuously, as was described in the previous paragraph.

7. Boundaries and boundary values

Remarkably, with a minimal amount of effort, and based on the mere definition of $F(w)$ (equation (40)), we were able to derive in [3] simple expressions for the location of the boundaries of the eigenvalue distribution and also for the boundary values of $\rho(r)$ in terms of the moments of $\tilde{\rho}(\sigma)$, which, I remind the reader, is the density of eigenvalues for a Hermitian matrix problem.

To this end, it is useful to rewrite our master formula (54) for $\gamma(r)$ as

$$wF(w) = \gamma \quad (72)$$

with

$$w = \frac{\gamma r^2}{\gamma - 1}. \quad (73)$$

We start with the outer edge $r = R_{\text{out}}$ (either in the disc phase or in the annular phase.) Near the outer edge $\gamma \to 1-$, and thus $w \to -\infty$. We therefore expand $F(w)$ in (72) in powers of $1/w$. As can be seen from (40), the coefficient of w^{-k-1} in this expansion is the kth moment

$$\langle\sigma^k\rangle = \int_0^\infty \tilde{\rho}(\sigma)\sigma^k \, \mathrm{d}\sigma \quad (74)$$

of $\tilde{\rho}(\sigma)$ (which is of course normalized to 1). For the class of models we are interested in here, all the moments $\langle\sigma^k\rangle$, $k \geqslant 0$ are finite. From the expanded form of (72) one can obtain, by a straightforward calculation [3], that

$$R_{\text{out}}^2 = \langle\sigma\rangle \quad (75)$$

and

$$\rho(R_{\text{out}}) = \frac{2R_{\text{out}}^2}{\langle\sigma^2\rangle - \langle\sigma\rangle^2}. \quad (76)$$

Thus, R_{out}^2 is simply the first moment of $\tilde{\rho}(\sigma)$, and the density $\rho(R_{\text{out}})$ is inversely proportional to the variance of σ.

For the $\tilde{\rho}(\sigma)$ under consideration here, $\langle\sigma^2\rangle$, and consequently $\rho(R_{\text{out}})$, are always finite. Outside the boundary $\rho(r)$ vanishes identically, of course, and thus, $\rho(r)$ always 'falls off a cliff' at the boundary, for all probability distributions of the form (20) with V polynomial. It would be thus interesting to study circularly invariant matrix ensembles $P(\phi^\dagger\phi)$ such that the eigenvalue distribution $\tilde{\rho}(\sigma)$ of $\phi^\dagger\phi$ has a finite $\langle\sigma\rangle$ but an infinite $\langle\sigma^2\rangle$. Then $\rho(R_{\text{out}})$ would vanish. This would naturally raise the question whether in such situations, $\rho(r)$ behaves universally near the edge (that is, if near the edge it vanishes like $(R_{\text{out}} - r)^\epsilon$ with ϵ being some universal exponent).

We now turn to the annular phase, and focus on the inner edge $r = R_{\text{in}}$ of the annulus. According to the discussion in the previous section (see equation (66) and the discussion above it), $a > 0$ in (57), and thus $F(w)$ is analytic in the domain $|w| < a$. We therefore expand $F(w)$ in (72) in powers of w. As can be seen from (40), the coefficient of w^{k-1} in this expansion is

the negative moment

$$\left\langle \frac{1}{\sigma^k} \right\rangle = \int_0^\infty \tilde{\rho}_{\text{annulus}}(\sigma) \frac{1}{\sigma^k} \, \mathrm{d}\sigma \tag{77}$$

of $\tilde{\rho}_{\text{annulus}}(\sigma)$. From (57) (or (66)) it is possible to see that all these negative moments of $\tilde{\rho}_{\text{annulus}}(\sigma)$ exist. From this expanded form of (72) one can obtain, again, by a straightforward calculation [3], that

$$\frac{1}{R_{\text{in}}^2} = \left\langle \frac{1}{\sigma} \right\rangle \tag{78}$$

and

$$\rho(R_{\text{in}}) = \frac{2R_{\text{in}}^{-6}}{\langle \sigma^{-2} \rangle - \langle \sigma^{-1} \rangle^2}. \tag{79}$$

Thus, R_{in}^{-2} is simply the σ^{-1} moment of $\tilde{\rho}_{\text{annulus}}(\sigma)$, and the density $\rho(R_{\text{in}})$ is inversely proportional to the variance of σ^{-1}. Since $\langle 1/\sigma^k \rangle$, $k = 1, 2$ are finite, $\rho_{\text{annulus}}(\sigma)$ jumps from zero (in the inner void of the annulus) to a finite value at the inner edge R_{in}. It can be shown [3], however, that when $a \to 0$, that is, in the annular to disc transition, $\langle 1/\sigma \rangle$ remains finite, but $\langle 1/\sigma^2 \rangle$ diverges like $1/\sqrt{a}$. Thus, from (78) we see that $R_{\text{inner}}(a = 0)$, the *critical* inner radius, is finite. The annulus starts up with a finite inner radius. Also, in this limit, we see from (79) that $\rho(R_{\text{in}})$ vanishes like $\sqrt{a}$. As we approach the annulus–disc transition, the discontinuity in $\rho(r)$ at the (finite) inner edge disappears.

We saw at the end of the previous section (see equations (65)–(68)) that $F(w)$ is continuous through the disc–annulus phase transition. Thus, our master formula $wF(w) = \gamma$ to determine $\gamma(r)$ (equation (72)) is also continuous through the transition. Consequently, $\rho(r) = (1/r)(\mathrm{d}\gamma/\mathrm{d}r)$ must remain continuous through the disc–annulus transition, and has (at the transition) the universal behaviour described in the previous paragraph.

Acknowledgments

It is a great pleasure to thank A Zee and R Scalettar for a fruitful and exciting collaboration. I also thank the organizers of the workshop for inviting me to present this work and for the very kind hospitality at the University of Stellenbosch during the conference.

References

[1] Feinberg J and Zee A 1997 *Nucl. Phys.* B **504** 579
[2] Feinberg J and Zee A 1997 *Nucl. Phys.* B **501** 643
[3] Feinberg J, Scalettar R and Zee A 2001 *J. Math. Phys.* **42** 5718
[4] Stephanov M A 1996 *Phys. Rev. Lett.* **76** 4472
Halasz M A, Jackson A D, Shrock R E, Stephanov M A and Verbaarschot J J M 1998 *Phys. Rev.* D **58** 096007
Verbaarschot J J M and Wettig T 2000 Random matrix theory and chiral symmetry in QCD *Ann. Rev. Nucl. Part. Sci.* **50** 343
Verbaarschot J J M 2004 QCD, chiral random matrix theory and integrability, lectures given at the *Les Houches Summer School on Applications of Random Matrices in Physics (Les Houches, June)* (*Preprint* hep-th/0502029)
Nowak M A 2001 *Lectures on Chiral Disorder in QCD (Proc. Cargese Summer School on QCD Perspectives on Hot and Dense Matter) (NATO Science Series II, Mathematics, Physics and Chemistry vol 87)* ed J-P Blaizot and E Iancu (Dordrecht: Kluwer)p 433–60 (*Preprint* hep-ph/0112296)
Janik R A, Nowak M A, Papp G and Zahed I 1998 Chiral random matrix models in QCD *Acta Phys. Pol.* B **29** 3957

Janik R A, Nowak M A, Papp G and Zahed I 1997 Various shades of blue's functions *Acta Phys. Pol.* B **28** 2949
[5] Janik R A, Nowak M A, Papp G, Wambach J and Zahed I 1997 *Phys. Rev.* E **55** 4100
Janik R A, Nowak M A, Papp G and Zahed I 1997 *Nucl. Phys.* B **501** 603
[6] Verbaarschot J J M, Weidenmüller H A and Zirnbauer M R 1985 Grassmann integration in stochastic quantum physics: the case of compound-nucleus scattering *Phys. Rep.* **129** 367
Fyodorov Y V and Sommers H-J 1997 Statistics of resonance poles, phase shifts and time delays in quantum chaotic scattering: random matrix approach for systems with broken time-reversal invariance *J. Math. Phys.* **38** 1918
Fyodorov Y V and Sommers H-J 2003 Random matrices close to Hermitian or unitary: overview of methods and results *J. Phys. A: Math. Gen.* **36** 3303 (special issue on 'random matrix theory')
Fyodorov Y V, Khoruzhenko B A and Sommers H-J 1997 *Phys. Rev. Lett.* **79** 557
Fyodorov Y V, Khoruzhenko B A and Sommers H-J 1997 *Phys. Lett.* A **226** 46
Fyodorov Y V, Khoruzhenko B A and Sommers H-J 1998 *Ann. Inst. Henri Poincaré* **68** 449
Fyodorov Y V 1999 Almost-Hermitean random matrices: applications to the theory of quantum chaotic scattering and beyond *Supersymmetry and Trace Formulas: Chaos and Disorder* ed I Lerner *et al* (Dordrecht/New York: Kluwer/Plenum) p 293
Gudowska-Nowak E, Papp G and Brickmann J 1997 *Preprint* cond-mat/9701187
[7] Haake F, Izrailev F, Lehmann N, Saher D and Sommers H J 1992 *Z. Phys.* B **88** 359
Sommers H J, Crisanti A, Sompolinsky H and Stein Y 1988 *Phys. Rev. Lett.* **60** 1895
[8] Hatano N and Nelson D R 1997 *Phys. Rev. Lett.* **77** 570
Hatano N and Nelson D R 1997 *Phys. Rev.* B **56** 8651
Hatano N and Nelson D R 1998 *Phys. Rev.* B **58** 8384
Hatano N 1998 Localization in non-Hermitean quantum mechanics and flux-line pinning in superconductors *Physica* A **254** 317 (*Preprint* cond-mat/9801283)
Zee A 1998 A non-Hermitean particle in a disordered world *Physica* A **254** 300 (*Preprint* cond-mat/9711114)
[9] Efetov K B 1997 *Phys. Rev. Lett.* **79** 491
Efetov K B 1997 *Phys. Rev.* B **56** 9630
[10] Kottos T 2005 Statistics of resonances and delay times in random media: beyond random matrix theory *J. Phys. A: Math. Gen.* **38** 10761 (special issue on 'aspects of quantum chaotic scattering') (*Preprint* cond-mat/0508173)
Rusek M, Mostowski J and Orlowski A 2000 *Phys. Rev.* A **61** 022704
[11] Chalker J T and Wang Z J 1997 *Phys. Rev. Lett.* **79** 1797
Chalker J T and Wang Z J 2000 *Phys. Rev.* E **61** 196
[12] Stephanov M A, Verbaarschot J J M and Wettig T 2001 Random matrices *Wiley Encyclopedia of Electrical and Electronics Engineering supplement 1* (New York: Wiley) (*Preprint* hep-ph/0509286)
[13] Brézin E, Itzykson C, Parisi G and Zuber J-B 1978 *Commun. Math. Phys.* **59** 35
[14] Ginibre J 1965 *J. Math. Phys.* **6** 440
[15] Bessis D, Itzykson C and Zuber J-B 1980 *Adv. Appl. Math.* **1** 109
For more recent references see, e.g., Brézin E and Zee A 1994 *Phys. Rev.* E **49** 2588
Zee A 2003 *Quantum Field Theory in a Nutshell* (Princeton, NJ: Princeton University Press) (chapter VII.4)
[16] Brézin E and Zee A 1993 *C. R. Acad. Sci.* **317** 735
Brézin E and Zinn-Justin J 1992 *Phys. Lett.* B **288** 54
Higuchi S, Itoh C, Nishigaki S and Sakai N 1993 *Phys. Lett.* B **318** 63
Higuchi S, Itoh C, Nishigaki S and Sakai N 1995 *Nucl. Phys.* B **434** 283
Higuchi S, Itoh C, Nishigaki S and Sakai N 1995 *Nucl. Phys.* B **441** 405 (erratum)
D'Anna J and Zee A 1996 *Phys. Rev.* E **53** 1399
Feinberg J and Zee A 1997 *J. Stat. Phys.* **87** 473
[17] There are many papers on chiral matrices. A partial list is as follows: Brézin E, Hikami S and Zee A 1995 *Phys. Rev.* E **51** 5442
Brézin E, Hikami S and Zee A 1996 *Nucl. Phys.* B **464** 411
Verbaarschot J J M 1994 *Nucl. Phys.* B **426** 559
Verbaarschot J J M and Zahed I 1993 *Phys. Rev. Lett.* **70** 3852
Cicuta G M and Montaldi E 1984 *Phys. Rev.* D **29** 1267
Barbieri A, Cicuta G M and Montaldi E 1984 *Nuovo Cimento* A **84** 173
Canali C M, Cicuta G M, Molinari L and Montaldi E 1986 *Nucl. Phys.* B **265** 485
Cicuta G M, Molinari L, Montaldi E and Riva F 1987 *J. Math. Phys.* **28** 1716
Slevin K and Nagao T 1993 *Phys. Rev. Lett.* **70** 635
Slevin K and Nagao T 1994 *Phys. Rev.* B **50** 2380

Nagao T and Slevin K 1993 *J. Math. Phys.* **34** 2075, 2317
Nagao T and Forrester P J 1995 *Nucl. Phys.* B **435** 401
Andreev A V, Simons B D and Taniguchi N 1994 *Nucl. Phys.* B **432** 485
Hikami S and Zee A 1995 *Nucl. Phys.* B **446** 337
Hikami S, Shirai M and Wegner F 1993 *Nucl. Phys.* B **408** 415
Hanna C B, Arovas D P, Mullen K and Girvin S M 1994 *Preprint* cond-mat 9412102
[18] Anderson A, Myers R C and Periwal V 1991 *Phys. Lett.* B **254** 89
Anderson A, Myers R C and Periwal V 1991 *Nucl. Phys.* B **360** 463
Myers R C and Periwal V 1991 *Nucl. Phys.* B **390** 716
Ambjørn J, Jurkiewicz J and Makeenko Yu M 1990 *Phys. Lett.* B **251** 517
Nishigaki S 1996 *Phys. Lett.* B **387** 139
[19] Ambjørn J 1990 *Quantization of Geometry (Les Houches)* ed J Dalibard *et al* (see section 4.3 and references therein)
[20] 't Hooft G 1974 *Nucl. Phys.* B **72** 461
[21] Coleman S 1995 *Aspects of Symmetry (Selected Erice Lectures)* (Cambridge: Cambridge University Press) chapter 8.3
[22] Brézin E and Neuberger H 1991 *Nucl. Phys.* B **350** 513
[23] Zee A 1996 *Nucl. Phys.* B **474** 726